Group theory of icosahedral virus capsid vibrations: a top-down approach

(Version published in Journal of Theoretical Biology, **256** 4, (2009), 607- 624)

Kasper Peeters[1] **and Anne Taormina**[2]

[1] Institute for Theoretical Physics
Utrecht University
P.O. Box 80.195
3508 TD Utrecht
The Netherlands

[2] Department for Mathematical Sciences
Durham University
South Road
Durham DH1 3LE
United Kingdom

kasper.peeters@aei.mpg.de, anne.taormina@durham.ac.uk

Abstract:

We explore the use of a top-down approach to analyse the dynamics of icosahedral virus capsids and complement the information obtained from bottom-up studies of viral vibrations available in the literature. A normal mode analysis based on protein association energies is used to study the frequency spectrum, in which we reveal a universal plateau of low-frequency modes shared by a large class of Caspar-Klug capsids. These modes break icosahedral symmetry and are potentially relevant to the genome release mechanism. We comment on the role of viral tiling theory in such dynamical considerations.

Contents

1. Introduction and summary

Viral protein capsids exhibit a surprising amount of symmetry, the consequences of which are still far from fully explored. A major step forward in the understanding of the *static* properties of icosahedral virus capsids was the introduction of viral tiling theory [1]. Viral tiling makes use of non-crystallographic Coxeter groups [2] to predict the distribution of proteins on the faces of the capsid. It fills an important hole in the Caspar-Klug classification [3], as it captures all known icosahedral viruses, in particular those with an all-pentamer structure which so far defied classification. There is also evidence that viral tiling plays a role in describing the clustering of proteins into building blocks during capsid assembly [4]. The use of group theory thus provides important insights not easily obtained otherwise (for a more extensive review of viral tiling, see [5]).

Apart from understanding static properties of viral capsids, one would also like to understand the influence of symmetries on the *dynamics* of these capsids. A first step in this direction was taken in [6], where it was shown how the presence of approximate inversion symmetry puts a clearly visible stamp on the spectrum of Raman and infrared vibrational modes. This analysis suggests that there may be various properties of capsid vibrations which have a rather simple explanation and do not require a great deal of knowledge about the detailed structure of the capsid.

A lot of effort has been dedicated to the study of protein vibrations in the last forty years, and more recently on viral capsid vibrations. The main goal so far has been to understand whether conformational changes, which are of the utmost importance for the proper function of these systems, occur in directions which overlap with some low-frequency normal modes of vibration. It is however also interesting to identify as precisely as possible the frequencies of viral vibrations which are susceptible to be targeted by near-infrared femtosecond laser pulses, in an attempt to mechanically destroy viral particles without damaging neighbouring tissues [7]. Despite decades of encouraging results, almost all relying on variations of the spring-mass model and high performance computing, the overall picture remains patchy and deserves to be revisited.

In order to determine the spectrum of vibrations of a viral capsid, one must, in principle, analyse a complicated interaction potential between a large number of atoms (of the order of 10^5 or 10^6). This potential depends on bond lengths and angles, the minimum of which describes the static configuration of the capsid. Small vibrations are described by a harmonic potential around the equilibrium situation. The normal modes of the associated force matrix yield the small fluctuation spectrum of the capsid. Computer simulations of this type have led to useful insight, about the relation between normal and swollen forms of virus capsids [8, 9] for instance, albeit at considerable computational cost (for recent reviews see [10, 11]).

It is clear that for the lowest-frequency modes, which are characterised by slow motions with large amplitudes, many degrees of freedom are not relevant. Various coarse-graining procedures have therefore been proposed, among which the popular rotation-translation block method [9, 12] and the cluster normal mode analysis [13] (see [14] for a review). These have shown that, indeed, some aspects of capsid dynamics can be understood from far fewer degrees of freedom than those present in all-atom simulations. Many of these developments were driven by the requirement to bring down the substantial computational cost of all-atom simulations. However, even if computational cost is not an issue, it is important to reduce the number of degrees of freedom in order to gain insight into the *systematics* of capsid dynamics.

Instead of using the "bottom-up" approach described above, in which one starts from the all-atom structural information and attempts to coarse-grain from there, we suggest here that useful complementary information can be obtained by using a "top-down" approach. The starting point is a very minimalistic description of the capsid, in which entire protein chains are modelled by point masses. These masses interact with each other through springs, whose strength is determined by the association energies of the protein chains. The relative values for the association energies are taken from VIPERdb [15]. This approach is similar in spirit to the one underlying viral tiling, in the sense that we attempt to see how much we can learn from such an extreme caricature before introducing more degrees of freedom or more complicated interactions.

Although this drastic simplification cannot possibly be expected to describe any but the lowest frequency modes, we will show here that it *does* uncover the systematics behind

several vibration patterns which were observed in all-atom simulations before, but for which a simple explanation remained lacking. Further insight will be obtained by decomposing the spectrum into irreducible representations of the icosahedral group (a method previously used for all-atom simulations [16, 17]). The power of the "top-down" approach is thus to make it much easier to see the common aspects of vibrational patterns among a large class of capsids. Moreover, it allows us to systematically think about the qualitative aspects of *modifying* vibrational modes (by changing bonds for example) without having to rely on the output of large-scale computer simulations.

We begin in section 2 with an overview of the role played by group theory in the analysis of normal modes of vibration of systems exhibiting some symmetry, and we state clearly the dynamical hypotheses on which our modelling relies. Section 3 presents the low-frequency spectra for eight viral capsids at various T-numbers, namely Satellite Tobacco Mosaic, Rice Yellow Mottle, Tomato Bushy Stunt, Cowpea Chlorotic Mottle, Polio, MS2, Hong Kong '97 and Simian 40. It emerges from the results obtained that the low-frequency spectrum of all these stable[1] capsids – bar the last one, which is in a different class, having an all-pentamer structure – possess 24 near-zero normal modes of vibration which always fall in the same set of non-singlet irreducible representations of the icosahedral group. The first singlet representation, which is associated with a fully symmetric mode, always appears higher up in the spectrum, in accordance with the expectation that such a motion requires more energy to develop. We argue in section 4 that the presence of 24 near-zero-modes in the spectrum of viral capsids is deeply rooted in the fact that the latter exhibit icosahedral symmetry. This is done by examining the clusters of protein chains which make up a face of the icosahedral structure, and then linking the clusters from different faces in the minimal possible way to obtain a stable structure. The mathematical argument is illustrated by considering a hypothetical and very simple viral capsid, where a single protein occupies the centre of each icosahedral face[2], and all proteins are linked to form a dodecahedron, as was first presented in [18]. More realistic capsids are then considered in the light of the group theoretical properties discovered in the simple case. We also offer some remarks on the role of tiling theory in our approach to virus dynamics. Finally, we conclude with some thoughts for future work in this direction.

2. Viral capsid vibrations: a group theory perspective

2.1 Symmetry-based organisation of normal modes of vibration

The most direct evidence of icosahedral symmetry in a large class of viruses is the experimental observation that the proteins, which form an almost spherical protective shell or capsid for the DNA or RNA material, do cluster in groups of five (pentamers) around twelve equidistant disclinations corresponding to the vertices of an icosahedron, and in groups of six (hexamers) around points at the intersection of the capsid with global 3-fold and/or local 6-fold symmetry axes of the same icosahedron. This symmetry is at the core of the Caspar-Klug classification of icosahedral viruses, which exploits well-chosen coordinate axes on a

[1] Some capsids are not stable when one considers the association energies given in VIPERdb alone. In these cases, extra theoretical bonds have been added to stabilise the capsid, and their occurrence is clearly stated in the text.

[2] One may think of this approximation as being a limiting case for a $T = 1$ capsid whose three protein chains per face are averaged by their centre of mass.

2-dimensional planar hexagonal lattice to label viruses according to the number of capsid proteins they exhibit [3].

Recent theoretical work by Twarock [1] corroborates the existence of an intrinsic icosahedral symmetry, even in the case of all-pentamer capsids which appear in the polyoma- and papilloma-viridae families. Apart from twelve pentamers located around the global 5-fold symmetry axes of the icosahedron, more pentamers are organised around local 5-fold symmetry axes, whose existence is deeply rooted in the mathematics of the icosahedral group H_3 [3]. A combination of ideas, inspired by the affinisation of this non-crystallographic Coxeter group [19] as well as the theory of quasicrystals and Penrose tilings, has led to interesting new insights on viral capsid assembly and genome organisation [4, 20–22].

Symmetry arguments have proven very powerful in a variety of scientific contexts, and it is natural to exploit the icosahedral symmetry to the full in attempts to understand dynamical properties of viruses, such as their vibration patterns. The techniques, applicable to any complex system whose N building blocks (considered as N point masses) are invariant under the action of a symmetry group G [4], date back to Wigner [23] and have been applied in chemistry as early as 1934 [24]. Since then, the frequencies of vibration of the most standard chemical molecules, and of some fullerene structures, have been analysed with group theory methods in the approximation where the potential is quadratic.

Remarkably, many qualitative features of the normal modes of vibration depend mainly on the symmetry of the system, and not on the potential chosen. For instance, it is quite straightforward to decompose the motions of a 'G-invariant N-atom molecule' into irreducible representations of the symmetry group G without reference to the harmonic potential chosen to model interactions between the atoms. Such decomposition provides further insight into the universality of certain motions, and we emphasise that it does *not* imply that we only keep the modes which respect the symmetry G. The potential is required to be G-invariant, but the solutions to the equations of motion, i.e. the vibration modes, can break this symmetry.

The 'molecule' we are interested in is a viral capsid, whose 'atoms' are its N capsid proteins, each of them approximated by a point mass located at its centre of mass. The symmetry group G is the icosahedral group H_3 if the capsid exhibits a centre of inversion, or its 60-dimensional proper rotation subgroup $\mathcal{I}$ otherwise. We remark here that Nature does not seem to favour viruses whose capsid proteins have a centre of inversion at the atomic level. Nevertheless, in coarse-grained approximations, including the one used here, the configuration of the N point masses chosen to model the capsid may be very close to having a centre of inversion. Consequences of this remark were explored in [6] in the context of viral tiling theory.

The decomposition of the vibrational modes of such a capsid into irreducible representations of the group H_3 (resp. $\mathcal{I}$) is standard and is sketched in appendix A.2. For the proper rotation subgroup $\mathcal{I}$ of the icosahedral group, the generic decomposition of the displacement representation (which encodes the action of $\mathcal{I}$ on the displacements of the N protein-point masses from their equilibrium position due to vibrations) reads,

$$\Gamma^{3N}_{\text{displ}} = \frac{3\,N}{60}\Big[\Gamma^1 + 3\Gamma^3 + 3\Gamma^{3'} + 4\Gamma^4 + 5\Gamma^5\Big]\,, \tag{2.1}$$

where all representations Γ on the r.h.s. are irreducible, and the numerical superscripts refer to the dimension of the representations. The decomposition (2.1) does not contain any

[3] The notation $\mathcal{I}_h$ is used in the chemistry literature.

[4] We will call such a system 'a G-invariant, N-atom molecule' although this is obviously an abuse of language.

information about the force matrix, introduced in appendix A.1, and hence does not provide any insight into the geometry of the vibration modes. In order to obtain that information, while preserving the group theoretical structure, one must turn to a block diagonalisation of the force matrix, as illustrated in appendix A.2 in the case of the ammonia molecule. Such a detailed analysis reveals that, in accordance with the decomposition (2.1), the capsid undergoes vibrational motions which can be expressed in terms of a linear combination of independent normal modes, some of them degenerate, in the sense that they all have the same frequency and are transformed into each other under the action of $\mathcal{I}$. The expression (2.1) encodes how the $3N$ normal modes of vibration are organised in subsets (irreducible representations) such that all normal modes within a given subset have the same frequency and transform into each other under the group action. The number of elements in a subset is given by the dimension on the corresponding irreducible representation.

Not all $3N$ normal modes of vibration of our viral capsids are 'genuine'. Indeed, in this 3-dimensional problem, three degrees of freedom correspond to the translations, and three to the rotations of the capsid as a whole. These zero-modes belong to two Γ^3 irreducible representations and are usually discarded from dynamical considerations as they are trivial.

Further standard group theory considerations provide an easy method to pin down which are, among the normal modes described in (2.1), those one could in principle detect using Raman and infrared spectroscopy. The former relies on a physical phenomenon induced by sending a beam of frequency ν on a molecule or capsid: some radiation is scattered, whose frequency is the incident frequency ν shifted by $\pm\nu_s$. The electric field $\vec{E}$ carried by the incident beam induces a dipole moment $\vec{\mu}$ in the molecule or capsid, such that

$$\mu_i = \alpha_{ij} E_j, \qquad i,j = 1,2,3\,, \tag{2.2}$$

and the polarisability tensor α_{ij} transforms under the symmetry group of the molecule as the six quadratic expressions $x_i x_j$, where $x_i, i = 1,2,3$ are the coordinates of a point in space. For the group $\mathcal{I}$, the quadratic expression $x_1^2 + x_2^2 + x_3^2$ transforms as a singlet (and thus belongs to Γ^1) while the other five independent quadratic expressions belong to Γ^5. From (2.1), one thus concludes that our N protein-point mass capsid has $3N/60$ non-degenerate[5] and $N/4$ five-fold degenerate Raman active modes of vibration. On the other hand, infrared active modes belong to Γ^3 irreducible representations (see [25] for instance), and therefore the capsid possesses $(9N/60 - 2)$ three-fold degenerate infrared active modes. Such information might become useful when experiments will be sensitive enough to measure very low-frequency modes in macro-biomolecular assemblies.

2.2 Overview of existing models

Normal mode analysis (NMA) has been quite successful in its attempts to describe the conformational changes in a variety of proteins [26–31], and it proves a useful tool in the study of the dynamics of large macro-biomolecular assemblies, in particular viruses. In this context, the idea is to verify whether various experimentally observed conformations of a given viral capsid could be inferred from each other by arguing that conformational changes occur in directions which maximally overlap with those of a few low-frequency normal modes of the capsid [8, 17, 32, 33].

The method, however, has limitations. One should keep in perspective that strictly speaking, biologically significant low-frequency motions are typically not vibrational, due to the

[5]The number N of capsid proteins of Caspar-Klug viruses is a multiple of 60.

damping influence of the environment. Furthermore, NMA assumes the existence of a single well potential whose minimum is a given stable configuration of the viral particle studied, overlooking the possibility of neighbouring multi-minima of energy. Also, the harmonic approximation to the single-well potential most analyses consider is only valid if the particle undergoes small motions, and this does not lend itself to an accurate description of the observed conformational changes. Nevertheless, NMA provides dynamical data which are consistent with experimental results, especially on proteins [34–37], and is supported by a recent statistical study [38]. It thus seems reasonable to continue to use the method, provided the results are interpreted in the light of the caveats above.

Viruses are much more complex structures than proteins, and the biggest challenge remains the choice, within the NMA framework, of a potential which captures the physics of capsid vibrations whilst taking into account a reduced number of degrees of freedom to enable practical calculations. Many NMA applied to viruses implement variations of the simple Elastic Network Model proposed a decade ago [39], in which the atoms are taken as point masses connected by springs modelling interatomic forces, provided the distance between them is smaller than a given cutoff parameter. Simplified versions include the restriction to C^α-atoms only, the approximation in which each residue is considered as a point mass, or where even larger domains within the constituent coat proteins are treated as rigid blocks [8].

The elastic potential in all analyses above has two major drawbacks: it does not discriminate between strong and weak bonds since it depends on a single spring constant, and it uses the rather crude technique of increasing the distance cutoff to resolve capsid instabilities. Consequently, the frequency spectra have much less structure than one would expect in reality, and in particular fail to reproduce areas of rigidity and flexibility of the capsid satisfactorily. This phenomenon is illustrated in appendix B, where a system of 8 atoms is studied from two different perspectives. On the one hand, the system is viewed as two proteins consisting of four atoms each, with interactions characterised by two different bond strengths. On the other hand, the system of 8 atoms is subjected to a Tirion potential with varying cutoff distance but using only one bond strength. The former is analysed within the rotations-translations of blocks method (RTB) [12] and provides a spectrum with enhanced structure, when set against the Tirion-based analysis.

In order to obtain more accurate spectra, the authors in [40–42] implement a bond-cutoff method. An elastic network whose representatives are $N\,C^\alpha$- atoms is set up such that four consecutive C^α-atoms are connected via springs, introducing $3N-6$ constraints in the system. This backbone modelling provides stability of proteins with a less intricate elastic network than the one obtained via the distance-cutoff method. Further springs with different spring constants are added to model the various types of chemical interactions (disulfide bonds, hydrogen bonds, salt-bridges and van der Waals forces) within each protein. The proposed model reproduces conformational changes better than the conventional distance-cutoff simulations. Adapting this model to viruses would certainly be enlightening, but remains a computational challenge at present. Our model is close in spirit to the above, but takes fewer degrees of freedom into consideration. It should be seen as a first step towards an implementation of the programme in [40–42] for viral capsids.

2.3 Dynamical hypotheses in our model

We model capsids "top-down", that is, by starting from as few degrees of freedom as possible. A minimal set consists of the centre-of-mass positions of the protein chains, as well as sufficient bonds to make the capsid stable. For the equilibrium positions of the proteins we make use

of the data in VIPERdb [15]. The equilibrium positions of the protein chains are assumed to respect icosahedral symmetry.

The inter-protein forces away from equilibrium are approximated by a harmonic potential, but we allow for the spring constants to be different for every bond. Denoting the spring constant between protein n and m by κ_{mn}, we thus have

$$V = \sum_{\substack{m<n \\ m,n=1}}^{N} \frac{1}{2}\kappa_{mn}\Big(|\vec{x}_m - \vec{x}_n| - |\vec{x}^{\,0}_m - \vec{x}^{\,0}_n|\Big)^2 . \tag{2.3}$$

Here $\vec{x}_m$ denotes the actual position of the protein and $\vec{x}^{\,0}_m$ its equilibrium position. For small deviations from equilibrium, we can expand the potential as

$$V = \sum_{m=1}^{N} \frac{\partial V}{\partial x^i_m}\bigg|_{x=x^0} (x^i_m - x^{0i}_m) + \sum_{m,n=1}^{N} \frac{1}{2}\,\frac{\partial^2 V}{\partial x^i_m x^j_n}\bigg|_{x=x_0} (x^i_m - x^{0i}_m)(x^j_n - x^{0j}_n) + \dots . \tag{2.4}$$

As there is only very little spread in the values of the protein masses in virus capsids, we will normalise them all to one, thereby absorbing the overall mass into the spring constants κ_{mn}. The equations of motion for the deviations from equilibrium then become

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(x^i_m - x^{0i}_m\big) + F^{ij}_{mn}\big(x^j_n - x^{0j}_n\big) = 0\,, \qquad i,j = 1,2,3\,, \tag{2.5}$$

where the force matrix is obtained as the second derivative of the potential with respect to the positions, evaluated at the equilibrium positions. Explicitly, using the potential (2.3), one finds

$$F^{ij}_{mn} = \frac{\partial^2 V}{\partial x^i_m \partial x^j_n}\bigg|_{x=x^0} = \begin{cases} \displaystyle\sum_{p\neq m} \kappa_{mp}\frac{(x_m - x_p)^i (x_m - x_p)^j}{(x_m - x_p)^2}\bigg|_{x=x^0} & \text{if } m = n\,, \\ \displaystyle -\kappa_{mn}\,\frac{(x_m - x_n)^i (x_m - x_n)^j}{(x_m - x_n)^2}\bigg|_{x=x^0} & \text{otherwise}\,. \end{cases} \tag{2.6}$$

It is convenient to store all degrees of freedom in one vector, i.e. to introduce a vector $\vec{q}$ with components given by $q^{3m-3+i} = r^i_m \equiv x^i_m - x^{0i}_m$. In terms of this vector the equation of motion reads

$$\ddot{\vec{q}} + F\vec{q} = 0\,, \tag{2.7}$$

and the frequencies of the normal modes are given by the eigenvalues of the matrix F, as reviewed in A.1. Note that this matrix depends explicitly on the equilibrium positions of the masses.

The determination of the frequencies is therefore straightforward in principle. However, given the fact that one is faced with highly complex systems, the rapidly increasing number of degrees of freedom and the shape of the potential put severe constraints on actual calculations. Although results based on a truly phenomenological potential are still out of reach, our model nevertheless paves the way to more realistic situations. It introduces a drastic coarse-graining but on the other hand, provides a hierarchy between bonds, dictated by the relative values of association energies listed in VIPERdb. It also explores the conditions for iso-staticity of the viral capsids, in the light of the icosahedral symmetry inherent to the system. In other words, the model is simple enough to enable the identification of which type of bonds are crucial or not for the stability of the capsids, as will be developed in section 4.

Beforehand, we calculate in section 3 the low-frequency spectra of a variety of Caspar-Klug capsids within the technical framework summarised in the present section, and highlight the remarkable feature announced in the Introduction, namely the presence of a very low-frequency plateau of twenty-four normal modes in all Caspar-Klug viruses studied. We also analyse SV40, an all-pentamer capsid which falls out of the Caspar-Klug classification, and note that the nature of the plateau is slightly different. A proper understanding of this phenomenon requires further investigation and is beyond the scope of this paper.

3. Results for specific virus capsids

In order to explore the consequences of our dynamical hypotheses we have analysed a variety of virus capsids. Their Caspar-Klug T-numbers range from 1 to 7, and we have paid particular attention to $T = 3$ capsids in order to assess the role of viral tiling theory in shaping vibrational spectra. We have indeed chosen our $T = 3$ capsids so that the 3 tiling types (triangle, rhomb and kite) compatible with the structure of 12 pentamers and 20 hexamers are represented. Table 1 summarises our choice of capsids. We now discuss each capsid separately and highlight various interesting features as we go along.

name	abbr.	pdb code	T	approx. centre of inversion	tiling	zero modes
Satellite Tobacco Mosaic	STMV	1A34	1	no	triangle	6
Rice Yellow Mottle	RYMV	1F2N	3	no	triangle	6
Tomato Bushy Stunt	TBSV	2TBV	3	no	rhomb	30
Cowpea Chlorotic Mottle	CCMV	1CWP	3	no	rhomb	90
Polio	Polio	2PLV	3	no	kite	6
MS2	MS2	2MS2	3	yes	rhomb	90
Hong Kong '97	HK97	2FTE	$7l$	no	rhomb	6
Simian 40	SV40	1SVA	$7d$	yes	rhomb/kite	6

Table 1: Summary of the viruses analysed in this section, together with some of their fundamental properties. If the number of zero modes is larger than 6 additional bonds, over and above those given in VIPERdb, have to be added to stabilise the capsid.

3.1 Satellite tobacco mosaic virus

The STMV virus capsid is one of the simplest capsids to analyse since it has $T = 1$ and only 60 capsid proteins. An all-atom molecular dynamics simulation for STMV was reported on in [43]. One of the main conclusions of their analysis is that the empty capsid (i.e. without RNA content) is unstable. The instability manifests itself in the behaviour of two non-adjacent faces around a 5-fold symmetry axes, which sink into the interior [43]. Although a molecular dynamics simulation of this type cannot reach the very lowest frequency modes because the maximal evolution time is limited, we can still use this analysis to draw some general conclusions about the low-frequency spectrum. An important result is that the proteins remain essentially undeformed. This yields support to one of our dynamical hypotheses,

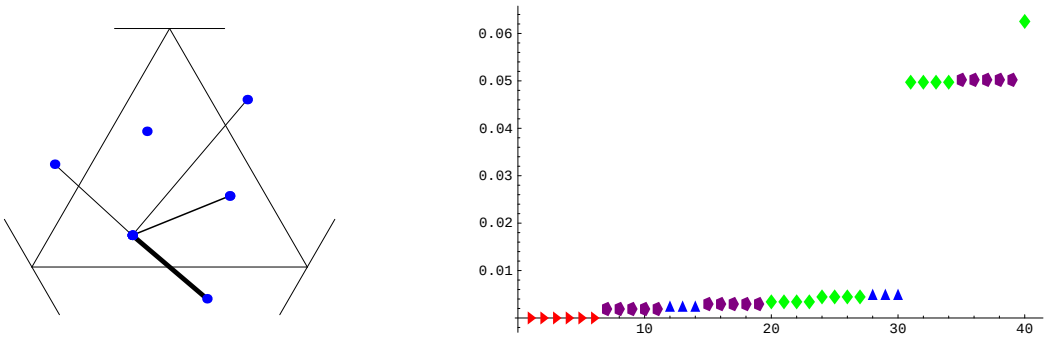

Figure 1: The STMV protein positions and the inter-protein bonds (left). The longest bond in the figure is the weak A1-A8 bond. Displayed on the right is the spectrum of the 40 lowest-frequency normal modes of STMV (up to an overall normalisation). Clearly visible are the 6 trivial zero-frequency modes as well as the 24 low-frequency modes. The modes $\triangleright$ (resp. $\triangle$) belong to 3-dimensional irreducible representations Γ^3 (resp. $\Gamma^{3'}$) of the icosahedral group. The diamond (resp. pentagon) modes belong to 4 (resp. 5)-dimensional irreducible representations. The x-axis labels the normal modes while the y-axis gives the wave numbers up to an overall normalisation.

namely that the individual atoms from which the proteins are built can be grouped together for an adequate analysis of the low-frequency modes. The collapse should thus be part of the low-frequency spectrum, as captured by a coarse-grained model.

In our simplified model the capsid is actually stable, but it has 24 normal modes with relatively low frequency. The bond structure, obtained from the association energies as described earlier, is displayed in figure 1. The resulting low-frequency spectrum is displayed there as well, with modes marked according to their representation content. The most significant feature of this spectrum is the appearance of a low-frequency plateau of 24 modes, separated by a large gap from the remainder of the spectrum. The representation content of the plateau is

$$2\Gamma^{3'} + 2\Gamma^4 + 2\Gamma^5 \,. \tag{3.1}$$

Within our model, it is easy to verify that the height of this plateau is related to the strength of the long A1-A8 bond. In the data provided by VIPERdb, this bond has a strength of only about 10% of the strongest bond present. By removing the A1-A8 bond, the 24 low-frequency plateau modes come down to zero frequency.

If one focuses only on the structure on each individual icosahedral face, one observes that the three protein chains form a relatively rigid triangle, which has 6 zero-modes (three rotations and three translations). Without the A1-A8 bond, there are 3 edge-crossing bonds per edge, as can be reconstructed from figure 1 by using the 2, 3 and 5-fold symmetries of the icosahedron. Since each bond induces one constraint in the system, naive counting suggest that there should be $6 \cdot 20 - 3 \cdot 30 = 30$ zero-modes in this case, as the icosahedron has 20 faces and 30 edges. We will see how this type of reasoning provides insight in many other virus capsids in section 4.2.

3.2 Rice yellow, tomato bushy stunt and cowpea chlorotic mottle virus

The RYMV, CCMV and TBSV capsids share many features, some of which are obvious from their common $T = 3$ structure, while others are less manifest. One of the most manifest differences is their viral tiling structure. While RYMV has a triangle tiling, both TBSV and

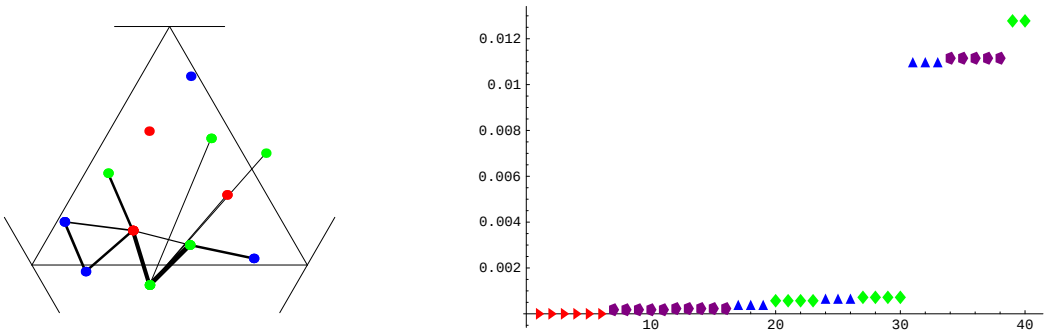

Figure 2: The protein positions and inter-protein bonds for RYMV (left), as well as the 40 lowest-lying normal mode frequencies (right). Again, there are 6 trivial zero-modes and 24 low-frequency modes, after which the frequencies go up rapidly.

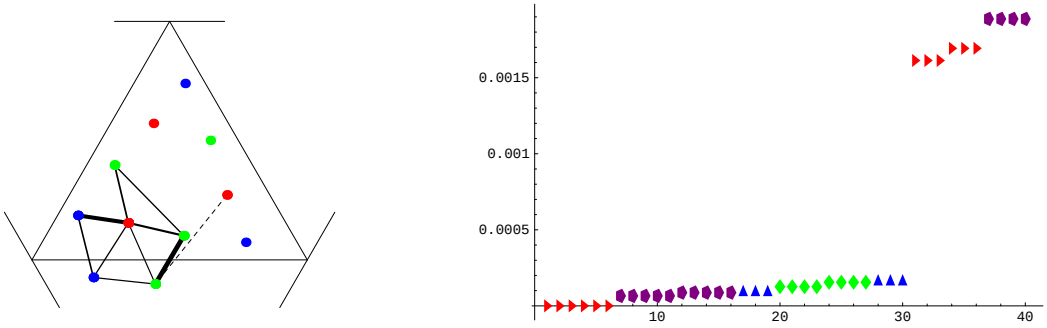

Figure 3: The protein positions and inter-protein bonds for TBSV (left), as well as the 40 lowest-lying normal mode frequencies (right). The bond indicated with a dashed line is a very weak bond which was added by hand to make the capsid stable.

CCMV are tiled by rhombs. The difference in tiling leads to a difference in dominant bond structure, but as we will see, this turns out to be of relatively little importance for the low-frequency spectrum. A somewhat related $T = 3$ virus, Polio, has an additional protein chain hidden slightly inside the main capsid. This kite-tiled virus will be discussed separately.

Let us start with RYMV, which has a triangle tiling. Its spectrum was computed using RTB techniques in [9] by making use of a Tirion potential with C^α atoms as fundamental degrees of freedom. Their analysis, however, is focused on the icosahedrically symmetric (non-degenerate, singlet) modes, while quite a few non-symmetric (degenerate) modes have lower frequency. In our model, employing VIPERdb association energies for the bonds, the capsid is stable and meaningful frequencies can thus be extracted. The spectrum of the first 40 modes is displayed in figure 2. Just as for the simpler STMV capsid discussed above, the most manifest feature is a plateau of 24 low-frequency modes (apart from the 6 trivial rotation and translation modes). This plateau in fact has precisely the same representation content as the one found for STMV, see (3.1).

The bond structure of RYMV exhibits, apart from "nearest neighbour" interactions, three long-range bonds which stretch out from the C-chains (see figure 2; one of these arms stretches from a (green) C-chain to a (red) B-chain and is hidden from view). These bonds are a consequence of the long tail of atoms which extends from the C-chain (see figure 4),

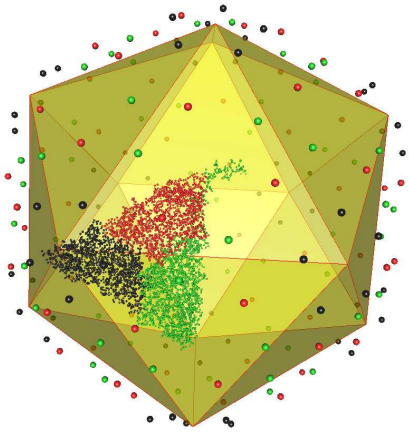

Figure 4: The capsid of RYMV. The large dots represent centre-of-mass positions of protein chains, while the small dots represent individual residues for three of the chains. Clearly visible is the long arm of the (green) C-chain which passes to the right of the (red) B-chain and stretches towards the top-right of the figure to establish the capsid-stabilising bonds.

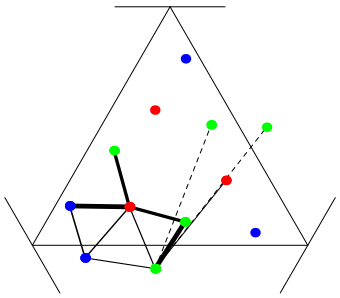

Figure 5: The protein positions and inter-protein bonds for CCMV as given in VIPERdb. The dashed bonds were added by hand to stabilise the capsid.

and turn out to be crucial for stability of the capsid. Artificially removing the weakest two introduces additional zero-modes which signal an instability against deformation. More precisely, what happens is that removal of these long-range bonds brings down the plateau of 24 low-frequency modes, turning them into zero-modes. Again, this is a perfect analogy with what happens for STMV.

Turning now to TBSV, we first of all note that the association energies given in VIPERdb are not sufficient to make the capsid stable. In fact, the TBSV spectrum has, in our model, a total of 30 zero-modes (including the trivial ones). By introducing one additional weak bond (indicated by a dotted line in figure 3), it turns out that 24 of these zero-modes get lifted. The resulting spectrum is again of the type we have seen before: a low-frequency 24-state plateau (3.1), followed by a gap.

Let us finally discuss CCMV. Its protein positions are similar to those of RYMV, but the bonds suggest a rhomb tiling instead of a triangular one. Unfortunately the association energies listed by VIPERdb are not sufficient to make the capsid stable, but we can add three weak long-range bonds similar to those in RYMV for stabilisation (see figure 5). Given the difference in structure of the strong bonds (as compared to RYMV), one might now naively

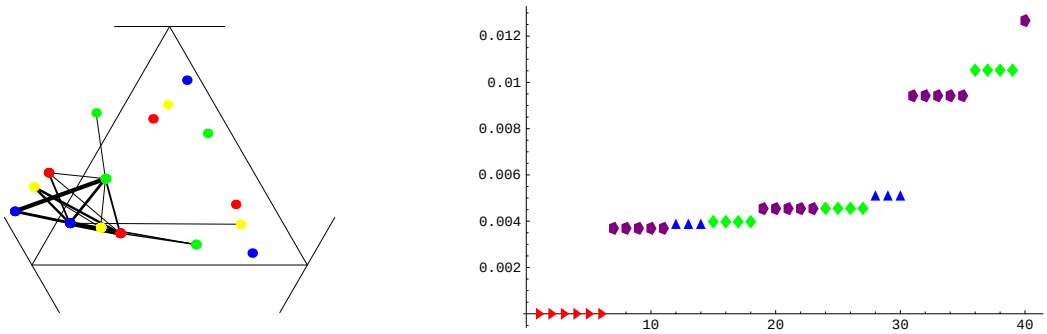

Figure 6: The protein positions and inter-protein bonds for the Polio virus (left), as well as the lowest-lying normal mode frequencies (right). Again, there are 6 trivial zero-modes and 24 low-frequency modes, after which the frequencies go up rapidly.

expect a rather different low-frequency spectrum. However, it turns out that CCMV again exhibits the by now familiar low-frequency 24-state plateau followed by a rather large gap (the size of which depends on the strength of the bonds which were added by hand). It is at present not easy to compare these results with other studies of the CCMV capsid. The analysis of [9] is based on a C^α-atom analysis using a Tirion potential. As we argue in appendix B, this potential has the tendency to smooth out sharp features of the spectrum. The fact that [9] find a singlet mode already after 17 non-singlet low-frequency modes may thus be due to this difference in the potential used. Dynamical aspects of CCMV have also been analysed by [44], who focused on the connectivity properties of the capsids rather than the precise form of the force matrix. Given the restrictions of the VIPERdb bonds, we will refrain from making a comparison with their results at this stage.

3.3 Polio virus

The polio virus is a pseudo $T = 3$ virus. It has four protein chains, one of which lies inside the main capsid. Ignoring this fourth chain, the tiling suggested by the remaining 180 protein positions is of kite type. The spectrum obtained from our model is displayed in figure 6. Once again, it exhibits a low-frequency plateau of 24 modes with the representation content (3.1). The lowest frequency modes, and in particular the characteristic motions of the proteins, have also been analysed to some extent in [17]. Of particular interest to us are their results on the representation content of the lowest frequency modes, which differs slightly from ours, and a visualisation of the motion of the atoms. The origin of this discrepancy remains unclear, as the latter paper is not very specific about the potential function used.

3.4 MS2

The inter-protein bonds for MS2 as given by VIPERdb are not sufficient to ensure stability of its capsid. For completeness we display the known bonds in figure 8. Although the number of zero-modes (90) is the same as for CCMV, we now no longer have access to a similar but stable capsid which we can use as a guideline to add stabilising bonds by hand (as we had for CCMV, where we used the similarity with RYMV). We therefore refrain from discussing the low-energy spectrum of MS2 here. It would be interesting to revisit the computation of the association energies for MS2, or alternatively get a better handle on the inter-protein bonds

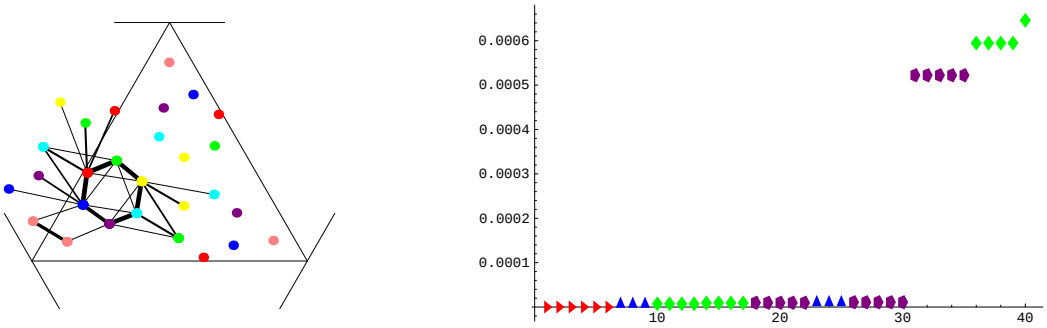

Figure 7: The protein positions and inter-protein bonds for HK97 (left) as well as the low-frequency spectrum (right).

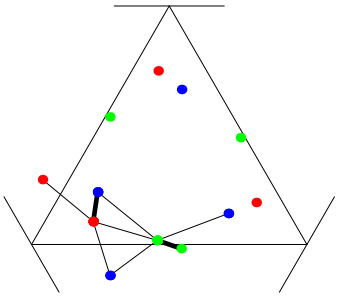

Figure 8: The protein positions and inter-protein bonds for MS2.

of this virus directly from experiment.

3.5 Hong Kong '97

The Hong Kong '97 virus has a $T = 7l$ capsid. Its hexamer rings are bound rather rigidly, as shown in figure 7. Once more we find that the low-frequency spectrum is dominated by a 24-state plateau, clearly separated from the remainder of the spectrum by a large frequency gap. The thickness of the faces is crucial for the stability of HK97, as a projection of the protein chains onto the hypothetical icosahedral faces leads to the appearance of three additional zero-modes. This yields a partial explanation for the extremely low frequency at which the plateau appears for HK97. We will return to the face thickness issue in section 4.2.

The spectrum of HK97 has previously been analysed in the Tirion approximation using RTB techniques by [9]. One of their key results is that the first two conformal modes correspond to protrusion of the hexamer and pentamer groups respectively. The structure of these singlet modes of HK97 has also been analysed by [45]. Our first singlet mode occurs well after the plateau, and has a structure which confirms these more elaborate computations. In particular, we see that the first singlet (depicted in figure 9) exhibits the same type of hexamer protrusion.

The structure of the 24-state low-frequency plateau which we observe is in agreement with the analysis of [33], who used an all-atom Tirion potential and a Lanczos eigenvalue solver to find that the first singlet occurs at position 31. However, we again note that the

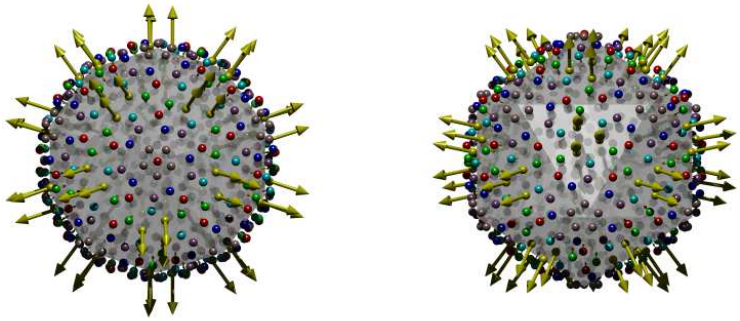

Figure 9: The first singlet mode of HK97, viewed along a 5-fold symmetry axis (left) and along a 3-fold axis (right). It occurs at position 67 with our particular bond choices, which is arguably still within the regime of the low-frequency approximation. This mode exhibits protrusion of the hexamer units (or rather, one of the chains of the hexamer), as also found in [45].

use of more crude approximations, such as in e.g. [9] or [13], gives a low-frequency structure that is different from ours.

3.6 Simian virus 40

Simian virus 40 has an all-pentamer capsid, which falls outside the Caspar-Klug classification scheme. It is captured, however, by viral tiling theory [1]. This particularity set aside, there is also a considerable difference in the bond structure as compared to other viruses we have analysed here. While HK97 exhibits a rather rigid structure around the edge of the hexamer, SV40 has a relatively weakly bound pentamer edge and much stronger bonds to other pentamers. See the left panel of figure 10 for details.

The difference between SV40 and all other capsids discussed so far is also clearly visible in the low-frequency spectrum: instead of the 24-state plateau which we have seen for all Caspar-Klug viruses, SV40 shows a much smoother structure. If a plateau is visible at all, it now contains 30 states, with representation content

$$2\Gamma^3 + 2\Gamma^{3'} + 2\Gamma^4 + 2\Gamma^5 \,. \tag{3.2}$$

We have so far not found any other virus with this low-frequency structure in its spectrum, which suggests that it is not as universal as the 24-state plateau observed for the Caspar-Klug family, but confirming this will require a more elaborate scan through the non-Caspar-Klug capsids.

4. Generic patterns in frequency spectra of viral capsids

4.1 The dodecahedron

In the previous section we presented the frequency spectra for a selection of icosahedral viral capsids, calculated within the mathematical set-up discussed in section 2. We will now argue that the distinctive low-frequency plateau of twenty-four normal modes, observed for all Caspar-Klug capsids, is in fact a direct consequence of a few long-range edge-crossing bonds. More precisely, we will show that, after an appropriate separation of the degrees of

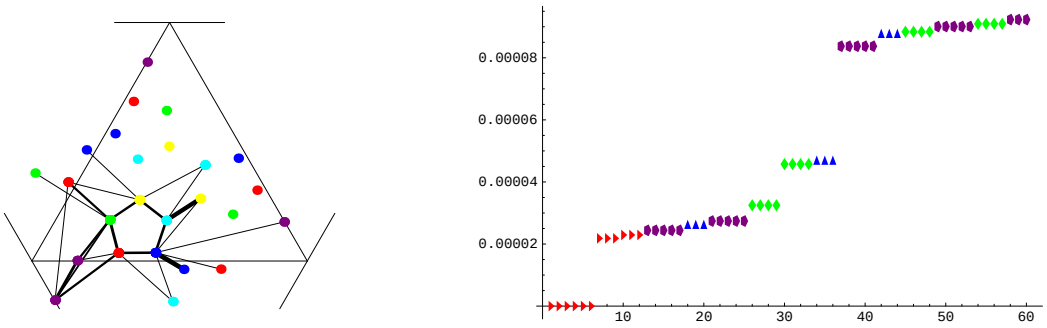

Figure 10: The protein positions and inter-protein bonds for SV40 (left) as well as the low-frequency spectrum (right).

freedom of the viral capsids, an explanation of the low-frequency plateau can be found in the vibrational spectrum of a simple dodecahedral spring-mass model [18].

In order to explain this, let us first focus on a caricature virus capsid, namely an hypothetical capsid with only one protein chain per icosahedral face, approximated by a point mass located at the centre of the face. Such a system possesses $3 \times 20 - 6 = 54$ degrees of freedom after subtraction of the trivial rotation and translation modes. We will also introduce a minimal set of bonds, obtained by connecting the nearest-neighbour proteins, and view the latter as the vertices of a polyhedron whose edges are the bonds. The resulting structure has thirty edges and is a *dodecahedron* dual to the icosahedron naturally associated with the capsid considered.

The thirty bonds reduce the number of degrees of freedom from 54 to 24, and these remaining degrees of freedom correspond to the 24 zero-modes which signal the instabilities of the dodecahedral cage. We now argue that the representation theory content of these zero-modes coincides exactly with that of the low-frequency modes of Caspar-Klug capsids. We first note that the dodecahedral spring-mass system considered here possesses a centre of inversion. Therefore, the relevant symmetry group is the full icosahedral group H_3 and the decomposition of the displacement representation of the dodecahedral capsid into a sum of irreducible representations of H_3 reads,

$$\Gamma_{\text{dode}}^{\text{disp},60} = \Gamma_+^1 \oplus \Gamma_+^3 \oplus 2\Gamma_-^3 \oplus \Gamma_+^{3'} \oplus 2\Gamma_-^{3'} \oplus 2\Gamma_+^4 \oplus 2\Gamma_-^4 \oplus 3\Gamma_+^5 \oplus 2\Gamma_-^5 \,. \tag{4.1}$$

The numerical superscripts refer to the dimension of the representations, while the $\pm$ signs differentiate between even and odd representations. This decomposition does not make use of any information stored in the force matrix, and therefore does not tell us which of these modes are zero-modes. In order to pin those down, we use the information encoded in the displacement representation of the icosahedron to which the dodecahedron constructed above is dual. First of all, the icosahedral vibration modes decompose as

$$\Gamma_{\text{ico}}^{\text{disp},36} = \Gamma_+^1 \oplus \Gamma_+^3 \oplus 2\Gamma_-^3 \oplus \Gamma_-^{3'} \oplus \Gamma_+^4 \oplus \Gamma_-^4 \oplus 2\Gamma_+^5 \oplus \Gamma_-^5 \,. \tag{4.2}$$

In this context of vibrations, the link between the dodecahedron and the icosahedron comes from considering the subset of motions of the dodecahedral system which are induced by the motion of the 12 vertices of an icosahedron. By "induced" we mean that the dodecahedron moves in such a way that its vertices are located at the centre of the deformed icosahedral

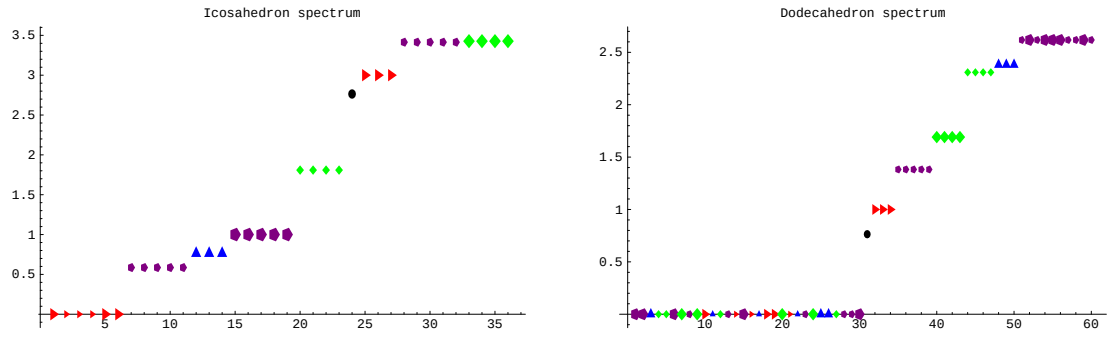

Figure 11: The spectrum of the icosahedron (left) and dodecahedron (right), decomposed in irreducible representations of the icosahedral group. Large symbols denote "−" representations, small symbols denote "+" representations. Black dots denote singlets, red triangles $\Gamma_{3^\pm}$, blue triangles $\Gamma_{3'^\pm}$, green squares $\Gamma_{4^\pm}$ and purple pentagons $\Gamma_{5^\pm}$. Note that the structure of the non-zero modes is the same for both spring-mass models.

faces, at all times. All icosahedral modes belonging to a particular irreducible representation must induce a linear combination of modes of the dodecahedral capsid which pertain to the same irreducible representation. Hence, provided that the vibrational modes of the icosahedron induce a non-vanishing component in the finite-frequency modes of the dodecahedron, we can conclude that the modes in $\Gamma^{\mathrm{disp},60}_{\mathrm{dode}}$ which are not contained in $\Gamma^{\mathrm{disp},36}_{\mathrm{ico}}$ must have vanishing frequency. Under this assumption, the zero-modes of the dodecahedron transform as

$$\Gamma^{\mathrm{disp}}_{\mathrm{dode,\ zero}} = \Gamma^{3'}_{+} + \Gamma^{3'}_{-} + \Gamma^{4}_{+} + \Gamma^{4}_{-} + \Gamma^{5}_{+} + \Gamma^{5}_{-}\,. \tag{4.3}$$

As a matter of fact, an explicit calculation of the normal modes confirms this assumption. The full frequency spectrum of the icosahedron and dodecahedron is displayed in figure 11 for reference.

We have visualised the twenty-four zero-modes in figure 12–14 and use the graphical representations to compare with the low-frequency modes of actual Caspar-Klug viruses.[6] In comparing with those of the Polio capsid in [17], one should recall that we perform a linearised analysis, and thus, any linear combination of eigenvectors with the same eigenvalue remains an eigenvector.

We end up this subsection with some insights in the multiplicity of the 24 zero-modes. Let us start with the 5-fold degenerate representations. Motion in the Γ_{5^+} representation is associated to squeezing in the direction of one of the six 5-fold axes of the icosahedron. A linear combination of all of them produces an icosahedrically symmetric motion, which is a singlet, so the degeneracy of the Γ_{5^+} modes is indeed five. Similarly, the modes in the Γ_{5^-} representation correspond to motion in which two opposite caps centred on a five-fold axis are rotated in opposite directions. Again, the linear sum produces a singlet, so the degeneracy is 5-fold. The Γ_{4^+} and Γ_{4^-} modes are related to the five cubes which are inscribed in the dodecahedral cage; again one linear combination transforms as a singlet (see also [46]). Generically, the dimensions of the irreducible representations are thus one lower than the multiplicity of modes with a similar geometrical pattern.

[6] Animations of the dodecahedron vibration modes are available at `http://biomaths.org/`.

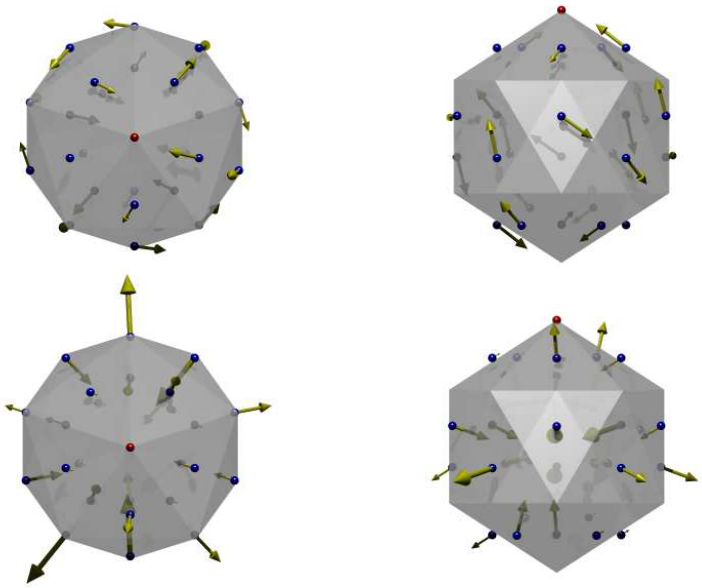

Figure 12: The zero-modes of the dodecahedron in the $\Gamma_{3'+}$ (top) and $\Gamma_{3'-}$ (bottom) representations.

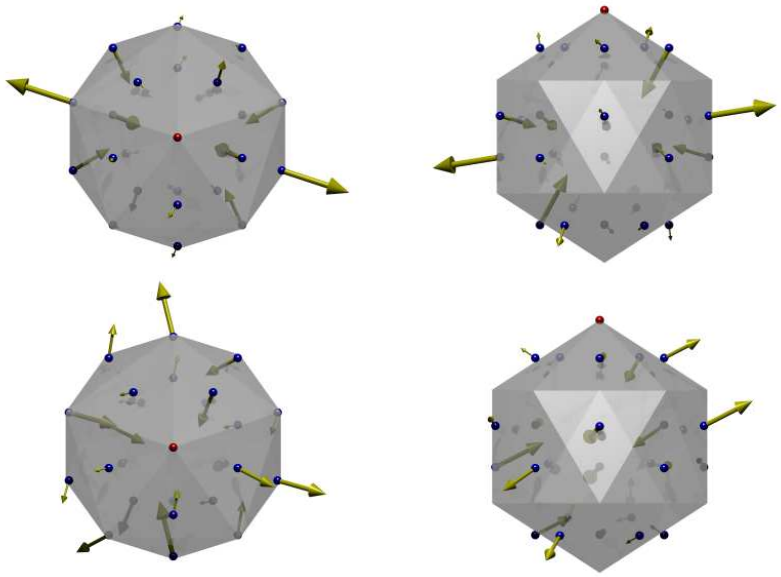

Figure 13: The zero-modes of the dodecahedron in the Γ_{4+} (top) and Γ_{4-} (bottom) representations.

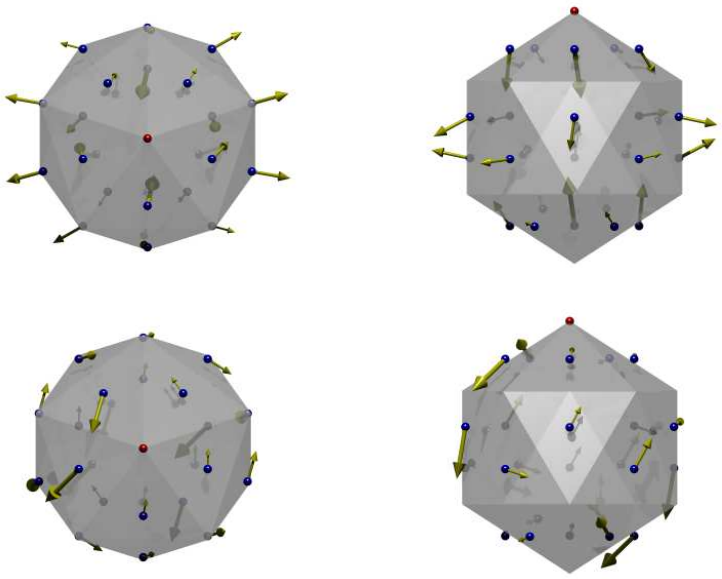

Figure 14: The zero-modes of the dodecahedron in the Γ_{5+} (top) and Γ_{5-} (bottom) representation.

4.2 Faces as building blocks

The analysis of the vibrations of the caricature virus capsid has revealed the existence of 24 zero-frequency modes. Despite its simplicity, this model can nevertheless be used to understand the low-frequency pattern of more complicated capsids, such as those analysed in section 3. The key ingredient here is to first analyse the vibration modes of the proteins on each icosahedral face (for the caricature capsid above, those vibration modes are just the three trivial translation modes).

Before we explain this logic in detail, let us summarise the situation for the various capsids which we have discussed in section 3. In figure 15 (resp. 16) we display the non-zero bonds of STMV, RYMV, Polio, HK97 (resp. SV40), restricted to those bonds which remain on a single icosahedral face. A crucial observation is that the faces are rarely stable: only in the case of STMV is there a sufficient number of inter-protein bonds to prevent the face from collapsing. The faces themselves thus exhibit zero-modes in their eigenspectrum. Additional bonds, which reach between proteins on different faces, are necessary for stability of the capsid as a whole and the faces in particular.

It is interesting to comment at this point on how zero modes arise. If all proteins associated with a given icosahedral face do actually lie in the plane of that face, "infinitesimal instabilities" may develop, as the forces that can be exerted by all bonds connected to a node lie in a plane. The force in the direction normal to the plane then vanishes at linear order, leading to a zero-mode of the force matrix. Figure 17 illustrates these infinitesimal instabilities in two and three dimensions. Generically, a fully connected network in the plane with n nodes has n zero-modes which move the individual nodes out of the plane (the z direction). Of these, one is a translation mode, and two others are rotations around the x- and y-axis respectively. Therefore, there are $n-3$ modes which are associated with infinitesimal instability. In the real world, the faces of the icosahedral capsids always have a finite thickness, because the distance of the protein chains to the centre of the capsid is not the same for

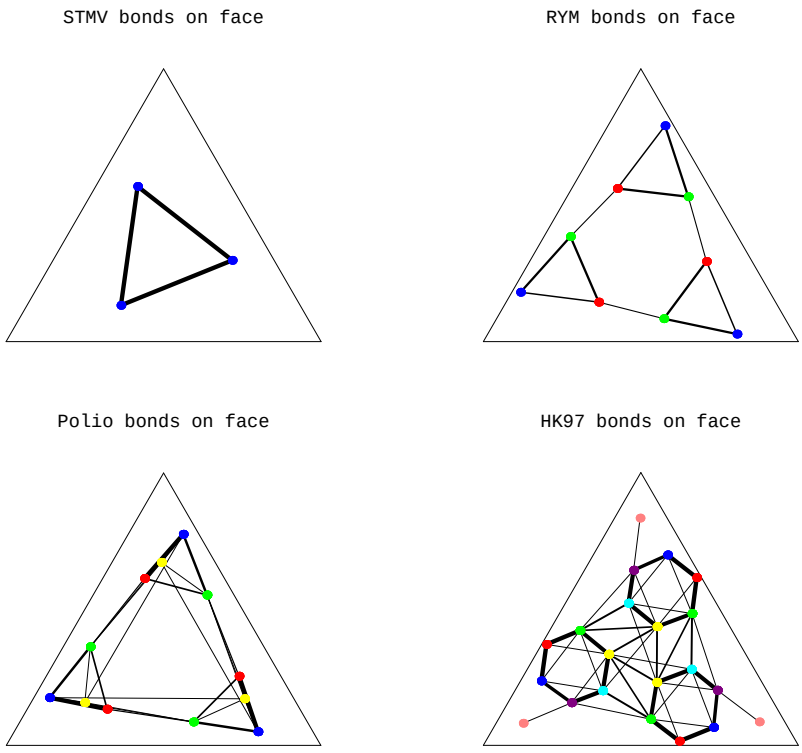

Figure 15: Structure of bonds on the faces of various virus capsids. These faces have 6(6), 15(15), 6(15) and 12(27) zero-modes respectively, where numbers in brackets include the infinitesimal instabilities which go away once the "thickness" of the face is taken into account.

all chains. As soon as this thickness is introduced, all infinitesimal instabilities disappear. Nevertheless, due to the fact that the faces are relatively thin, these lifted zero-modes will often have a small frequency.

Let us now turn to the general argument. A $T = n$ Caspar-Klug capsid has $3n$ proteins per face, and we will denote the number of links between these proteins with $3k$. If there were no links between proteins on different faces, one would expect a total of

$$N_0 = (9n - 3k) \cdot 20 = 60(3n - k) \tag{4.4}$$

capsid zero-modes. A stable capsid, which only has 3 rotational and 3 translational zero-modes, would thus require at least $60(3n - k) - 6$ independent constraints (i.e. bonds). How these bonds should be chosen remains an open mathematical problem. In two dimensions there exists an elegant algorithm to determine the stability of a network. This "pebble game" algorithm scales approximately linearly in the number of nodes [47]. The crucial ingredient which makes this possible is the Laman theorem [48], which allows one to find redundant bonds by analysing strict subgraphs of the full network. This theorem does not generalise to higher dimensions, and despite considerable effort an exact analogue of the pebble game in three dimensions is not known.

The VIPERdb database, however, provides us with various association energies for edge-crossing bonds which we can use to stabilise the capsids. By symmetry, these bonds come in multiples of 30 or 60 when one considers the capsid as a whole. The logic above suggests that, if we keep the strongest $e = 2(3n-k)-1$ bonds per edge, this would yield a capsid with

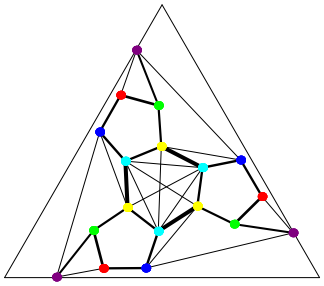

Figure 16: Structure of bonds on the face of SV40. Its faces have 18(21) zero-modes respectively (notation as in figure 15).

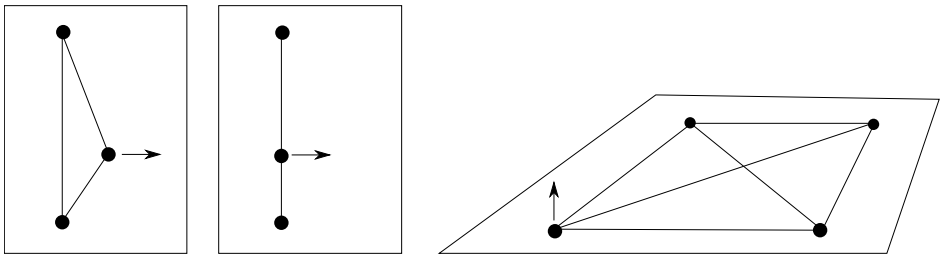

Figure 17: Infinitesimal instabilities arise when the restoring force vanishes at linear order. The 2d structure in figure a is stable, while the structures in b and c exhibit an infinitesimal instability.

24 non-trivial zero-modes. A careful analysis of the "on-face" bond structure given earlier confirms that this indeed is the case for all Caspar-Klug capsids which we have analysed. Here are some examples.

The STMV capsid, with $n = 1$, is the simplest case to analyse. It has a stable triangular face structure with $k = 1$, and from $2(3n - k) - 1 = 3$ we expect to need $e = 3$ edge-crossing bonds per edge. Looking at the left panel of figure 1 we observe that if we remove the weakest (longest) bond from the capsid, there are indeed 3 bonds per edge (the figure shows two of these bonds, of which one does not go through a two-fold axis and therefore really appears twice per edge after the icosahedral symmetries are taken into account). This reduced capsid is visualised in figure 18. An analysis of its spectrum indeed exhibits 24 non-trivial zero-modes in addition to the 6 translation and rotation modes, in agreement with the counting argument given above.

For the RYMV capsid, which has $n = 3$, if we discard the two weakest bonds (which are both edge-crossing) we end up with $k = 4$ on-face and $e = 9$ edge-crossing bonds, and 24 non-trivial zero-modes, in accordance with $2(3n-k)-1 = 9$. As discussed in section 3.2, the bonds that stretch from the C-chains are crucial for stability, and their weakness is responsible for the fact that the zero-modes are only slightly lifted. For HK97, which has $n = 7$, the situation is more subtle. Some of the weakest bonds on the faces of this capsid are actually weaker than some of the edge-crossing bonds. Consistently removing the weakest six bonds leads to a capsid with $k = 12$ and $e = 17$. The expression with $2(3n - k) - 1 = 17$ thus suggests that in this case there should be 24 non-trivial zero-modes, which we indeed observe. This type

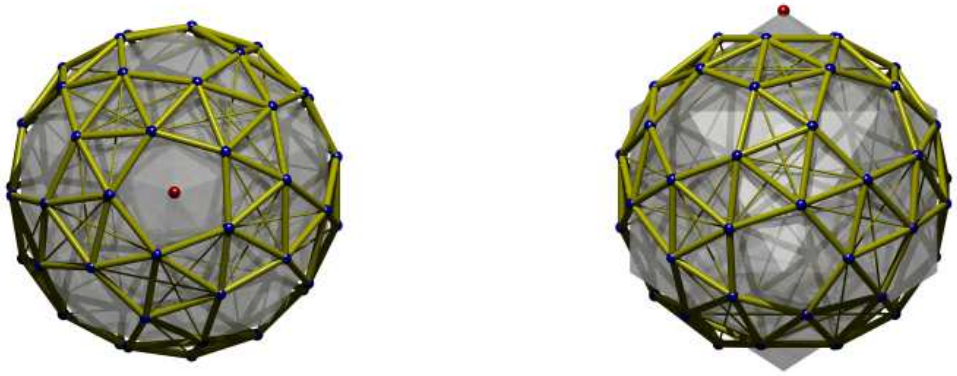

Figure 18: "Top" and "front" view of a geometric model of the capsid of STMV, including the inter-protein bonds. Dots denote the centre-of-mass positions of the protein chains, tubes denote bonds between them. This structure has 30 zero-modes without the A1-A8 bond (thin tubes); 24 of these are lifted by this weak bond.

of analysis can be extended to all the other Caspar-Klug viruses, with similar results.

4.3 Tiling independence

We have seen that, despite the considerable difference in the structure of Caspar-Klug capsids, the low-frequency spectrum exhibits a remarkable uniformity. A priori one might have expected that the way in which the strongest bonds are distributed over the capsid has a determining influence on the low-frequency vibration pattern.

In order to isolate the effects of variations in the bonding patterns from the effects of changing protein locations, we have analysed three hypothetical $T = 3$ capsids. These are all based on the protein positions of RYMV, but have different bond types corresponding to the three different $T = 3$ tilings: triangle, rhomb or kite. In figure 19 we display these hypothetical capsids, while in 20 we show the three possible tilings in their most ideal form. One can think of the capsids in figure 19 as small deformations of RYMV, CCMV and Polio (without 4th chain), respectively, as their protein positions are very similar. We have taken a hierarchy of bond strengths in which the strongest bonds are of unit strength, most other bonds are of strength 1/5, and the long C-chain arms are taken to be of strength α_1 and α_2 respectively. The positions of the bonds are based on those of RYMV, ignoring the weak long-range bond between the B- and C-chain.

An analysis of the low-frequency spectrum exhibits a 24-state plateau in all cases, provided that the bonds α_1 and α_2 are not simultaneously taken to be very small. Thus, the size of the gap is at least partially caused by one of the long-range bonds not being extremely small. When one of the $\alpha_{1,2}$ is set to zero, the 24 low-frequency modes come down to zero frequency. Interestingly, when both long-range bonds are very small, i.e. $\alpha_1 = \alpha_2 \ll 1$, the plateau extends to a total length of 64. In all three capsids, one then finds a state at position 71 which is a singlet conformal mode.

We thus conclude that the structure of the bonds is far less relevant than one would expect, and the spectrum is essentially determined by the few long-range edge-crossing bonds which make the capsid stable. This provides some support for the analysis in section 4.1, in which most results were obtained by analysing the displacement representation, which contains no

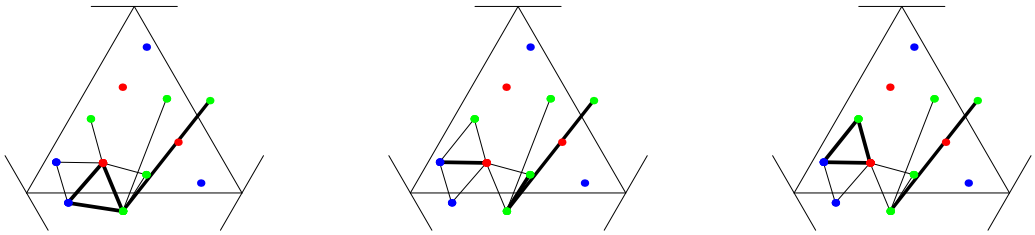

Figure 19: Three hypothetical virus capsids, corresponding to a triangle, rhomb and kite bond structure respectively.

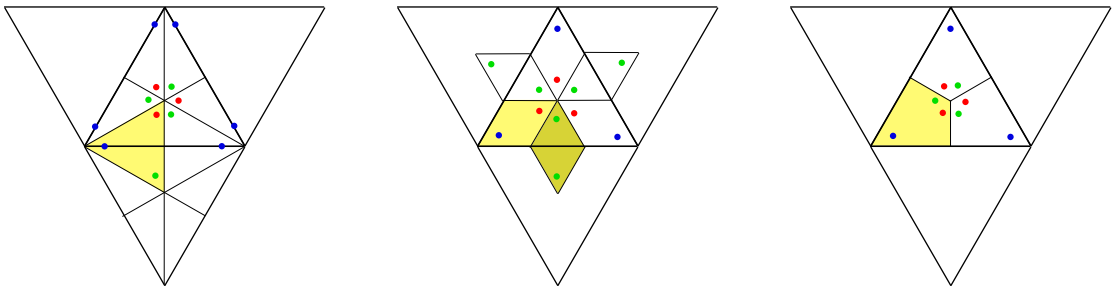

Figure 20: The three possible tilings for $T = 3$ capsids. The central triangle denotes one face of the icosahedron. The strongest inter-protein bonds are those between proteins on a single tile, i.e. on one of the yellow-coloured polygons. Bonds between equal-type chains are all *weak* in the Caspar-Klug and kite tilings, while some of them are *strong* in the rhomb tiling. The rhomb tiling is the only one without strong bonds between monomers on different faces.

information about bonds. However, there clearly are many mathematical questions which these observations raise, and we hope to return to some of these in a future publication.

5. Discussion and conclusions

Dynamical properties of viral capsids are undoubtedly key to our understanding of conformational changes, maturation and function of viruses. They may even offer clues on how to interfere mechanically and chemically with their assembly or, more generally, their replication cycle, with potentially important repercussions in the health sector.

In this paper, we have been concerned with slow, large amplitude vibrations of icosahedral capsids, which are thought to be crucial for the onset of conformational changes. The underlying techniques, which have been used for decades in the analysis of single protein vibrations, are in principle applicable in the present context, but the complexity of large biomolecular assemblies renders molecular dynamics simulations prohibitive in cost and computer time. A major tool in the study of viral capsid vibrations is Normal Mode Analysis (NMA), despite its limitations, briefly mentioned in section 2.2. The basic inputs of a NMA are the chosen level of coarse-graining for the system under consideration and the quadratic potential adopted. The few existing analyses are probing different aspects of virus vibrations, and are just too scarce to provide good ground for identifying general patterns across viruses.

Our approach has been to study low-frequency modes of vibration from a coarse-grained approximation which replaces each capsid protein by a point mass located at its centre of mass, while interactions are dictated by a spring-mass model with all masses being equal and spring constants reflecting the relative values of association energies listed in VIPERdb. Our analysis further assumes that the capsid is empty. Although our coarse-graining is certainly drastic, its combination with a quadratic potential sensitive to a variety of bond strengths has enabled us to detect universal properties of the frequency spectra, namely a low-frequency plateau of twenty-four states, which is present in all Caspar-Klug capsids analysed, i.e. in all capsids exhibiting twelve pentamers at the vertices of an icosahedron and hexamers respecting the icosahedral symmetry everywhere else, irrespectively of the tiling type (triangles, rhombs or kites).

Such signature is difficult to extract from most available molecular dynamics simulations we are aware of. In some cases (e.g. [8, 9]), we believe the reason is that the distance cut-off needed to stabilise capsids described by a Tirion potential smears the low-frequency spectrum and loses the plateau structures, a phenomenon illustrated in a toy model in appendix B. Another potential source of confusion comes from the fact that some analyses only keep normal modes of vibration which are invariant under the proper rotation subgroup $\mathcal{I}$ of the icosahedral group, while we argued in section 2.1 that although the potential must be invariant, the solutions to the equations of motion need not be. Yet, there is one analysis which is compatible with our results: in the all-atom simulation with cut-off 7.3Å of the HK97 capsid in [33], the first non-degenerate mode (a singlet of $\mathcal{I}$) appears at position 31 in the spectrum, in total agreement with our qualitative arguments on the number of low-frequency modes (6 trivial zero-modes and 24 very near zero-modes).

Let us, however, emphasise once more that we do not claim our approximations should be expected to be adequate for all virus capsids. It is conceivable that in some cases one cannot neglect contributions like bond-bending or angle-dependent forces, or cannot treat interactions sensibly in a perturbation series. Even if the degrees of freedom arise from point-like masses, as in the C_{60} fullerene for instance, our conclusion about the existence of a 24-state plateau does not necessarily hold. As a matter of fact, our approximations hint at a plateau of 84 low-frequency modes for C_{60}. A discussion of any gap after that plateau is likely to be invalidated by the fact that it would occur for too high frequencies in the spectrum. Experimentally one does not observe any noticable gap in the spectrum of C_{60}. Nevertheless, this observation for the fullerene does not invalidate our analysis for Caspar-Klug capsids, for which the number of very low-frequcny modes is much smaller.

Our top-down approach should therefore be seen as complementary to the all-atom simulations, and paves the way for a new generation of models which should address several shortcomings of the present coarse-graining methods. The most obvious extension which should be analysed is one in which angle-dependent forces are included. Within the harmonic approximation, an additional very interesting and deep mathematical question is how to minimise the number of bonds in a 3-dimensional spring-mass model so that the structure remains stable. In the absence of useful theorems, one can be guided by Nature and make use of the data on association energies in VIPERdb. However, there are several viral capsids for which interesting dynamics has been observed using the RTB method or all-atom normal mode computations, but which cannot currently be handled with the coarse-grained model proposed in this paper. This is due to the limited number of bonds listed in VIPERdb. Capsids for which the available bonds are insufficient, even after introduction by hand of one or two additional bonds, include Hepatitis B and MS2. In this light, it would be interesting

to revisit the computation of the association energies for these capsids. Another outstanding puzzle is a proper mathematical understanding of the different low-frequency plateau structure of SV40, in particular whether the 30 near zero-modes are a signature of all-pentamer capsids.

Finally, in view of the importance of conformational changes in viral capsids, which can only be fully characterised by going beyond the harmonic approximation, it is desirable to develop methods taking anharmonicity into account. For instance, one might consider generalising to viral capsids the Elastic Network Interpolation method developed in [40] to generate anharmonic pathways of macro-biomolecules.

Acknowledgements

We thank François Englert for an inspiring collaboration on some parts of the work reported here (see [18]). We also thank Peter Stockley, Roman Tuma and especially Reidun Twarock for discussions.

Appendix

A. Normal modes of vibration

A.1 Force matrix and normal coordinates

We consider a molecule with N atoms, which are approximated by point masses whose displacements from equilibrium due to vibrational motion are encoded in $3N$ mass-weighted coordinates $q_i(t), i = 1, .., N$, where t is time. We drop the explicit time dependence in what follows, and use the notation $\dot{q}_i = \mathrm{d}q_i/\mathrm{d}t$ and $\ddot{q}_i = \mathrm{d}^2 q_i/\mathrm{d}t^2$. We also use Einstein's summation convention, for instance, $\sum_{j=1}^{3N} A_{ij}\, v_j \equiv A_{ij}\, v_j$. The $3N$ classical equations of motion are

$$\ddot{q}_i + F_{ij} q_j = 0, \qquad i = 1, .., 3N\,, \tag{A.1}$$

where the kinetic energy is given by $T = \frac{1}{2}\dot{q}_i{}^2$ and the potential energy, Taylor expanded about its equilibrium position and normalised so that it vanishes at equilibrium, boils down, in the harmonic approximation, to the following expression

$$V = F_{ij} q_i q_j, \tag{A.2}$$

with the $3N \times 3N$ force matrix or Hessian given by,

$$F_{ij} = \left.\frac{\partial^2 V}{\partial q_i \partial q_j}\right|_{q_k = q_k^0}, \qquad \text{where } q_k^0 \text{ are the atoms' equilibrium positions.} \tag{A.3}$$

One is interested in changing basis from the general coordinates $\{q_i\}$ to a new set of coordinates $\{q'_k\}$ so that the set of differential equations (A.1) is equivalent to the set

$$\ddot{q}_k^{'} + \lambda_{(k)} q_k^{'} = 0, \qquad k = 1, .., 3N\,, \tag{A.4}$$

for some real values $\lambda_{(k)}$ to be determined. If the orthogonal matrix of change of basis is μ, i.e. $q'_k = \mu_{ki}\, q_i$, one must have, after constructing $3N$ linear combinations of (A.1) with coefficients C_{ki},

$$C_{ki}\ddot{q}_i + C_{ki} F_{ij} q_j = 0 = \ddot{q}_k^{'} + \lambda_{(k)} q_k^{'} \tag{A.5}$$

where $C_{ki} = \mu_{ki}$ and $C_{ki} F_{ij} = \lambda_{(k)} \mu_{kj}$. This implies that one must find $3N$ sets of $3N$ coefficients C_{kj} such that

$$(F_{ij} - \lambda_{(k)} \delta_{ij})\, C_{ki} = 0. \tag{A.6}$$

It is clear from (A.6) that, unless all coefficients are zero, the $\lambda_{(k)}$ are eigenvalues of the force matrix. For each such eigenvalue (fixed k), the equation (A.6) allows to determine $3N$ coefficients C_{ki}. The $3N$ normal coordinates $q_k^{'} = C_{ki} q_i$ are solutions of the differential equations (A.4), so they read,

$$q_k^{'} = A_k\, \cos(\lambda_{(k)}^{1/2}\, t + \epsilon_{(k)}), \qquad k = 1, .., 3N\,, \tag{A.7}$$

where $\epsilon_{(k)}$ are phases and A_k are constants. Reverting to the original system of coordinates,

$$q_i = C_{ij}\, q_j^{'} = C_{ij}\, A_j\, \cos(\lambda_{(j)}^{1/2}\, t + \epsilon_{(j)}), \qquad i = 1, .., 3N\,, \tag{A.8}$$

and choosing all but one constant A_k to be zero yields the i^{th} normal mode of vibration

$$q_i = C_{ik}\, A_k\, \cos(\lambda_{(k)}^{1/2}\, t + \epsilon_{(k)}), \qquad \text{no summation on } k \tag{A.9}$$

of frequency $\nu_{(k)} = \lambda_{(k)}^{1/2}/2\pi$.

A.2 Group theory patterns of normal modes of vibration

In this appendix we will briefly recall the method of decomposing vibration modes of a G-invariant N-atom molecule into irreducible representations of the symmetry group G (for more details see e.g. [49]). The goal is to find a new basis $\vec{q}^{\,\prime}$ for the atoms' positions, such that the action of G on $\vec{q}^{\,\prime}$ decomposes into a sum of irreducible representations. The action of the group G on $\vec{q}^{\,\prime}$ is called the *displacement representation*, and is simply given by the tensor product

$$\Gamma^{\rm disp}(g) = P(g) \otimes R(g)\,, \qquad g \in G\,. \tag{A.10}$$

Equivalently, in terms of the components, we have $(\Gamma^{\rm disp})_{3m-3+i,3n-3+j} = P(g)_{mn} \otimes R^{ij}(g)$, using the notation introduced above (2.7). The displacement representation matrices do not depend on the positions of the masses, but only on the number of atoms.

We denote with U the matrix which achieves the coordinate transformation $\vec{q}^{\,\prime} = U\vec{q}$ such that it block-diagonalises the displacement generators,

$$\Gamma'^{\,\rm disp}(g) := U^{-1}\Gamma^{\rm disp}(g)U = \oplus_p n_p \Gamma^p(g)\,. \tag{A.11}$$

Here $\Gamma^p(g)$ are the generators of the symmetry group G in the p-th representation. For every symmetry generator of the molecule which leaves the potential invariant, we necessarily have

$$\Gamma^{\rm disp}(g)\, F = F\, \Gamma^{\rm disp}(g), \tag{A.12}$$

where F is the force matrix introduced in appendix A.1. By Schur's lemma, the block diagonal form of the right-hand side of (A.11),

$$\Gamma'^{\,\rm disp}(g) = \begin{pmatrix} \underbrace{\begin{matrix} \Gamma^{p_1} & & \\ & \ddots & \\ & & \Gamma^{p_1} \end{matrix}}_{n_{p_1}\times n_{p_1}} & & \\ & \underbrace{\begin{matrix} \Gamma^{p_2} & & \\ & \ddots & \\ & & \Gamma^{p_2} \end{matrix}}_{n_{p_2}\times n_{p_2}} & \\ & & \ddots \end{pmatrix} \tag{A.13}$$

(in which all Γ^p are $d_p \times d_p$ matrices) thus implies a similar block-diagonal form of the U-transformed force matrix $F' = U^{-1}FU$,

$$F' = \begin{pmatrix} \underbrace{\begin{pmatrix} F^{p_1}_{11} & \cdots & F^{p_1}_{1\,n_{p_1}} \\ \vdots & \ddots & \vdots \\ F^{p_1}_{n_{p_1}1} & \cdots & F^{p_1}_{n_{p_1}\,n_{p_1}} \end{pmatrix}}_{n_{p_1}\times n_{p_1}} \otimes \mathbb{1}_{d_{p_1}\times d_{p_1}} & & \\ & \underbrace{\begin{pmatrix} F^{p_2}_{11} & \cdots & F^{p_2}_{1\,n_{p_2}} \\ \vdots & \ddots & \vdots \\ F^{p_2}_{n_{p_2}1} & \cdots & F^{p_2}_{n_{p_2}\,n_{p_2}} \end{pmatrix}}_{n_{p_2}\times n_{p_2}} \otimes \mathbb{1}_{d_{p_2}\times d_{p_2}} & \\ & & \ddots \end{pmatrix}. \tag{A.14}$$

Every irreducible representation Γ^p on the right-hand side of (A.11) thus corresponds to a set of n_p eigenvalues of F, each occurring with multiplicity d_p.

The coordinate transformation matrix U itself can be found using projection operators. These are defined as

$$\big(\mathcal{P}^p_{mn}\big)_{ab} = \frac{d_p}{\dim G} \sum_{g\in G} \big(\Gamma^p(g)\big)^*_{nm} \big(\Gamma^{\rm disp}(g)\big)_{ab}\,, \qquad m,n = 1,\ldots,d_p\,, \quad a,b = 1,\ldots,3N\,. \tag{A.15}$$

(In the case of the icosahedral group the representations are all real and the complex conjugation symbol can be ignored). A crucial property of these operators is that they provide us with basis vectors transforming according to the representation Γ^p. To see this, let us transform a *row* of the matrix $\mathcal{P}^p_{mn}$ by acting on it with a group element h in the displacement representation,

$$\begin{aligned}\big(\mathcal{P}^p_{mn}\big)'_{ba} = \big(\Gamma^{\rm disp}(h)\big)_{ac}\big(\mathcal{P}^p_{mn}\big)_{bc} &= \frac{d_p}{\dim G} \sum_{g} \big(\Gamma^p(g)\big)^*_{nm} \big(\Gamma^{\rm disp}(g)\,\Gamma^{\rm disp}(h^{-1})\big)_{ba} \\ &= \frac{d_p}{\dim G} \sum_{g'=gh^{-1}} \big(\Gamma^p(g')\big)^*_{nk} \big(\Gamma^p(h)\big)^*_{km} \big(\Gamma^{\rm disp}(g')\big)_{ba} \\ &= \big(\Gamma^p(h)\big)^*_{km} \big(\mathcal{P}^p_{kn}\big)_{ba}\,.\end{aligned} \tag{A.16}$$

This shows that the rows of the matrices $\mathcal{P}^p_{mn}$ transform into each other according to the p-th representation. In order to construct a transformation matrix which takes us to the symmetry-adapted basis, we thus have to construct the span of all row vectors in all $\mathcal{P}^p_{mn}$ matrices (for fixed p and all m,n). In fact, the normalised row-vectors of the matrices $\mathcal{P}^p_{mm}$, with $m = 1\ldots d_p$, together already span an n_p-dimensional subspace transforming as Γ^p. The matrix U^T is now constructed by taking an orthonormal basis in each of these subspaces, and then using these basis vectors to populate the rows of U^T. The matrix U then also block-diagonalises the force matrix F. Note once more that U only depends on the number of capsid proteins, not on their precise positions. For more details we refer the reader to e.g. [49].

If one is only interested in the multiplicities of the various irreducible representations, the U matrix is not needed and it is more convenient to use the character formula

$$n_p = \frac{1}{\dim G} \sum_{g\in G} \chi^{*\,\rm disp}(g)\,\chi^p(g)\,. \tag{A.17}$$

Here the character of a group element g in a given representation is obtained by computing the trace of the associated matrix, $\chi(g) = \mathrm{Tr}\,\big(R(g)\big)$. For the displacement representation (A.10) the character becomes

$$\chi^{\rm disp}(g) = \mathrm{Tr}\,\big(P(g)\big)\cdot \mathrm{Tr}\,\big(R(g)\big) = \pm(\#\text{ of proteins unmoved by } g)\cdot(1+2\cos\theta)\,, \tag{A.18}$$

where θ is the angle of the rotation associated with g, and the minus sign is taken whenever g involves an inversion. The characters of irreducible representations of finite groups are widely available in the literature, and are reproduced in appendix C for the icosahedral group H_3. This fixes the multiplicities n_p and can be used as a check on the intermediate step of the calculation of the projection operators.

Let us, to conclude, illustrate the technique described above for the case of the ammonia molecule (see figure 21), whose point group is C_{3v} and whose displacement representation matrices can be shown to satisfy

$$\Gamma^{\mathrm{disp}\,'}(g) = U^{-1}\,\Gamma^{\mathrm{disp}}(g)\,U = 3\,\Gamma^{1}(g) + \Gamma^{1'}(g) + 4\Gamma^{2}(g), \quad \forall g \in C_{3v}. \tag{A.19}$$

Hence the multiplicities are $n_1 = 3$, $n_2 = 1$ and $n_3 = 4$. The matrix $F^{'}$ can be partitioned in nine blocks of dimensions $n_i \times n_j, i, j = 1, 2, 3$ to match the structure of the displacement representation (A.19), and use the following notations

$$F^{'} = \left(\begin{array}{cccc|c|cccc} F_{11}^{'11} & F_{12}^{'11} & F_{13}^{'11} & & F_{11}^{'12} & F_{11}^{'13} & F_{12}^{'13} & F_{13}^{'13} & F_{14}^{'13} \\ F_{21}^{'11} & F_{22}^{'11} & F_{23}^{'11} & & F_{21}^{'12} & F_{21}^{'13} & F_{22}^{'13} & F_{23}^{'13} & F_{24}^{'13} \\ F_{31}^{'11} & F_{32}^{'11} & F_{33}^{'11} & & F_{31}^{'12} & F_{31}^{'13} & F_{32}^{'13} & F_{33}^{'13} & F_{34}^{'13} \\ \hline F_{11}^{'21} & F_{12}^{'21} & F_{13}^{'21} & & F_{11}^{'22} & F_{11}^{'23} & F_{12}^{'23} & F_{13}^{'23} & F_{14}^{'23} \\ \hline F_{11}^{'31} & F_{12}^{'31} & F_{13}^{'31} & & F_{11}^{'32} & F_{11}^{'33} & F_{12}^{'33} & F_{13}^{'33} & F_{14}^{'33} \\ F_{21}^{'31} & F_{22}^{'31} & F_{23}^{'31} & & F_{21}^{'32} & F_{21}^{'33} & F_{22}^{'33} & F_{23}^{'33} & F_{24}^{'33} \\ F_{31}^{'31} & F_{32}^{'31} & F_{33}^{'31} & & F_{31}^{'32} & F_{31}^{'33} & F_{32}^{'33} & F_{33}^{'33} & F_{34}^{'33} \\ F_{41}^{'31} & F_{42}^{'31} & F_{43}^{'31} & & F_{41}^{'32} & F_{41}^{'33} & F_{42}^{'33} & F_{43}^{'33} & F_{44}^{'33} \end{array}\right)$$

In the above, the expressions $F_{\alpha\beta}^{'pq}, \alpha = 1, .., n_p; \beta = 1, .., n_q; p, q = 1, 2, 3$ are matrices of dimension $d_p \times d_q$.

The equality $\Gamma^{\mathrm{disp}\,'}(g)F' = F'\Gamma^{\mathrm{disp}\,'}(g) \quad \forall g \in C_{3v}$ is equivalent to

$$\Gamma^{p}(g)F_{\alpha\beta}^{'pq} = F_{\alpha\beta}^{'pq}\Gamma^{q}(g), \quad \forall g \in C_{3v}, \tag{A.20}$$

for $\alpha = 1, .., n_p; \beta = 1, .., n_q; p, q = 1, 2, 3$. We are now in a position to make good use of Schur's Lemma. If $p \neq q$, the irreducible representations Γ^p and Γ^q are not equivalent, and this implies

$$F_{\alpha\beta}^{'pq} = \left\{\begin{array}{ll} 0, & \text{if } p \neq q \\ f_{\alpha\beta}^{p}\,\mathbf{I}_{d_p}, & \text{if } p = q \end{array}\right. \tag{A.21}$$

where $f_{\alpha\beta}^{p}$ are a set of $(n_p)^2$ complex numbers obeying $f_{\alpha\beta}^{p} = (f_{\beta\alpha}^{p})^*$ as F, and hence $F^{'}$ is

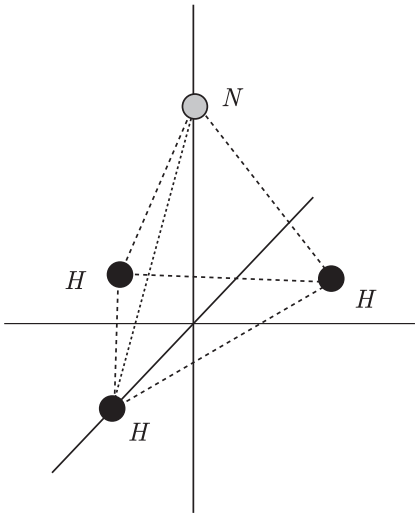

Figure 21: The structure of ammonia.

hermitian. So Schur's Lemma and the hermiticity of $F^{'}$ allow us to write

$$F^{'} = \left(\begin{array}{ccc|c|cccccccc} f^1_{11} & f^1_{12} & f^1_{13} & & & & & & & & & \\ (f^1_{12})^* & f^1_{22} & f^1_{23} & & & & & & & & & \\ (f^1_{13})^* & (f^1_{23})^* & f^1_{33} & & & & & & & & & \\ \hline & & & f^2_{11} & & & & & & & & \\ \hline & & & & f^3_{11} & 0 & f^3_{12} & 0 & f^3_{13} & 0 & f^3_{14} & 0 \\ & & & & 0 & f^3_{11} & 0 & f^3_{12} & 0 & f^3_{13} & 0 & f^3_{14} \\ & & & & (f^3_{12})^* & 0 & f^3_{22} & 0 & f^3_{23} & 0 & f^3_{24} & 0 \\ & & & & 0 & (f^3_{12})^* & 0 & f^3_{22} & 0 & f^3_{23} & 0 & f^3_{24} \\ & & & & (f^3_{13})^* & 0 & (f^3_{23})^* & 0 & f^3_{33} & 0 & f^3_{34} & 0 \\ & & & & 0 & (f^3_{13})^* & 0 & (f^3_{23})^* & 0 & f^3_{33} & 0 & f^3_{34} \\ & & & & (f^3_{14})^* & 0 & (f^3_{24})^* & 0 & (f^3_{34})^* & 0 & f^3_{44} & 0 \\ & & & & 0 & (f^3_{14})^* & 0 & (f^3_{24})^* & 0 & (f^3_{34})^* & 0 & f^3_{44} \end{array}\right)$$

where all entries not explicitly written are zero. After a further similarity transformation that reshuffles rows 7, 9 and 11, and then columns 7, 9 and 11, we arrive at (noticing also

that F and U are real)

$$F'' = \left(\begin{array}{ccc|c|cccc|cccc} f^1_{11} & f^1_{12} & f^1_{13} & & & & & & & & & \\ f^1_{12} & f^1_{22} & f^1_{23} & & & & & & & & & \\ f^1_{13} & f^1_{23} & f^1_{33} & & & & & & & & & \\ \hline & & & f^2_{11} & & & & & & & & \\ \hline & & & & f^3_{11} & f^3_{12} & f^3_{13} & f^3_{14} & & & & \\ & & & & f^3_{12} & f^3_{22} & f^3_{23} & f^3_{24} & & & & \\ & & & & f^3_{13} & f^3_{23} & f^3_{33} & f^3_{34} & & & & \\ & & & & f^3_{14} & f^3_{24} & f^3_{34} & f^3_{44} & & & & \\ \hline & & & & & & & & f^3_{11} & f^3_{12} & f^3_{13} & f^3_{14} \\ & & & & & & & & f^3_{12} & f^3_{22} & f^3_{23} & f^3_{24} \\ & & & & & & & & f^3_{13} & f^3_{23} & f^3_{33} & f^3_{34} \\ & & & & & & & & f^3_{14} & f^3_{24} & f^3_{34} & f^3_{44} \end{array}\right) \tag{A.22}$$

So we have managed to transform the force matrix F into an equivalent matrix, F'', which is block diagonal. It now remains to calculate the eigenvalues and the eigenvectors of the smaller size matrices appearing in the block diagonal form to obtain the normal modes of vibration. Note that each eigenvalue of F is associated with one irreducible representation Γ^p of C_{3v}, and each such eigenvalue is d_p-fold degenerate, where d_p is the dimension of Γ^p.

B. The Tirion and RTB approximations

In order to understand the approximations made with the Tirion potential and the rotations-translations of blocks (RTB) method, let us consider a simple example. This example models two protein chains, each individually bound together by strong covalent bonds, and mutually interacting through weak Van der Waals bonds. This is a caricature of the real world situation, but useful to understand the logic of these two approximations.

The model we will consider lives in two dimensions. We have two protein chains of 4 atoms each, positioned on the nodes of a square. These two chains interact weakly, as in figure 22.

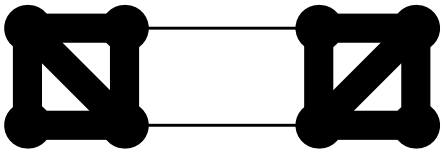

Figure 22: Two simplified protein chains whose mutual interaction is very weak compared to the interatomic forces within each protein. The thickness of the lines symbolises the strength of the bonds.

This "protein" has one zero-mode in addition to the three trivial ones (translation and rotation). This additional zero-mode corresponds to shearing motion of the central square.

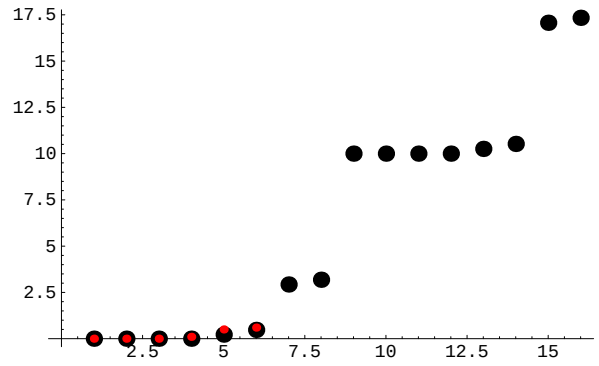

Figure 23: Exact frequency spectrum of the two protein chains' system.

In order to exhibit the consequences of a large hierarchy of interaction strengths, we will take the ratio of the strong to weak bonds to be 10. The force matrix of this model is computed as described earlier, keeping track of the different spring constants. The exact spectrum of this model is given by the black dots in the plot of figure 23.

In the "rotations-translations of blocks" (RTB) method, one computes the spectrum of the low-frequency modes by grouping together clusters (blocks) of atoms which form relatively rigid structures. By making these clusters completely rigid and keeping only their translation and rotation degrees of freedom, the total computational complexity is reduced. In our example above, we will use the two protein chains as blocks. There are two translations for each block, and one rotation, yielding a total of 6 coordinates. These are given by

$$\begin{aligned} t_x^1 &= \frac{1}{2}(e_x^1 + e_x^2 + e_x^7 + e_x^8)\,, \\ t_y^1 &= \frac{1}{2}(e_y^1 + e_y^2 + e_y^7 + e_y^8)\,, \qquad r^1 = \frac{1}{2}(e_x^1 + e_x^2 - e_x^7 - e_x^8)\,, \\ t_x^2 &= \frac{1}{2}(e_x^3 + e_x^4 + e_x^5 + e_x^6)\,, \qquad r^2 = \frac{1}{2}(e_x^3 + e_x^4 - e_x^5 - e_x^6)\,. \\ t_y^2 &= \frac{1}{2}(e_y^3 + e_y^4 + e_y^5 + e_y^6)\,, \end{aligned} \tag{B.1}$$

Note that the rotation coordinates are so simple because we only look at the linearised approximation. The resulting projection matrix P reads

$$P = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{B.2}$$

This projection matrix is orthogonal, i.e. it satisfies

$$P \cdot P^T = 1_{6\times 6}. \tag{B.3}$$

The force matrix for the reduced system now reads

$$P \cdot F \cdot P^T = \frac{1}{4} \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 11 & -1 \\ 0 & 0 & 0 & 0 & -1 & 11 \end{pmatrix} . \tag{B.4}$$

The eigenvalues of this matrix are

$$0,\ 0,\ 0,\ \frac{1}{2},\ \frac{5}{2},\ \frac{6}{2} . \tag{B.5}$$

These frequencies are, as expected [50], larger than the actual frequencies. They are depicted by the small red dots in the figure above, where a scaling by a factor of 1/5 was applied. We thus see that this method does a decent job at reproducing the low-frequency spectrum of the exact model, and correctly throws away all the high-frequency modes.

Now consider the Tirion approximation. In this approximation, we forget about all details of the bonds, and instead write down a potential which depends on only one overall spring constant (to be fixed by hand). Moreover, the potential depends on a cutoff radius. In figure 24 we show three configurations for three cutoff radii $R_c = 2.1$, 3.1 and 5.1. We

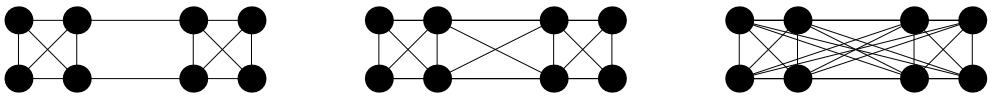

Figure 24: Visualisation of the Tirion bonds of the toy model protein for three cutoff radii..

arbitrarily fix the spring constant to be 3, so the plots of the spectrum should be read modulo an overall normalisation. For the three cutoff radii mentioned above, the spectra are presented in figure 25. From these plots it is obvious that this method can produce results

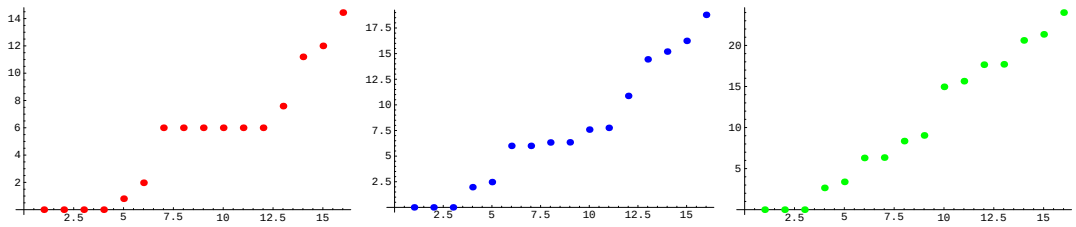

Figure 25: The exact frequency spectra for the Tirion potential with three different cutoffs, for the two protein chains' system.

which are both quantitatively and qualitatively wrong. In particular, it fails to reproduce the separation between high-frequency and low-frequency modes; instead, one is "supposed to know" where to stop trusting the Tirion spectrum. Note, however, that a sufficiently small cutoff can *still* change the zero-mode structure and introduce instabilities (the standard solution, see e.g. [51], is to simply increase the cutoff length until the capsid is stable, but this seems rather arbitrary).

C. Irreducible representations and characters of the icosahedral group

For completeness we list here the irreducible representations of the icosahedral group, based on those of [52]. There are ten irreducible representations, with generators labelled as $M_g^{i\pm}$, where $i = 1 \ldots 5$ (labelling the representation) and $g \in \{-, 2, 5\}$ (indicating the inversion element, the two-fold and the five-fold generator respectively). They thus come in pairs, in which the action of the inversion element in the '+' representation is trivial, while it inverts all coordinates in the '−' representation. That is,

$$M_-^{i+} = 1\,, \qquad M_-^{i-} = -1\,. \tag{C.1}$$

The form of the two-fold and the five-fold rotation generators is listed below. Here $\tau = \frac{1}{2}(1+\sqrt{5})$ denotes the golden ratio.

$$M_2^{1\pm} = 1\,, \qquad M_5^{1\pm} = 1\,,$$
$$M_2^{2\pm} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad M_5^{2\pm} = \frac{1}{2}\begin{pmatrix} \tau^{-1} & -\tau & 1 \\ \tau & 1 & \tau^{-1} \\ -1 & \tau^{-1} & \tau \end{pmatrix}, \tag{C.2a}$$

$$M_2^{3\pm} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad M_5^{3\pm} = \frac{1}{2}\begin{pmatrix} -\tau & -\tau^{-1} & 1 \\ \tau^{-1} & 1 & \tau \\ -1 & \tau & -\tau^{-1} \end{pmatrix},$$

$$M_2^{4\pm} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad M_5^{4\pm} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \tag{C.2b}$$

$$M_2^{5\pm} = \begin{pmatrix} 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}, \qquad M_5^{5\pm} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix}.$$

For the decomposition of the displacement representation into irreducible representations we frequently make use of the characters of the icosahedral group. These are listed in table 2, both for H_3 as well as its proper rotational subgroup $\mathcal{I}$.

conj. class	$\mathcal{C}(e)$	$\mathcal{C}(g_5)$	$\mathcal{C}(g_5^2)$	$\mathcal{C}(g_3)$	$\mathcal{C}(g_2)$	$\mathcal{C}(g_0)$	$\mathcal{C}(g_0g_5)$	$\mathcal{C}(g_0g_5^2)$	$\mathcal{C}(g_0g_3)$	$\mathcal{C}(g_0g_2)$
size	1	12	12	20	15	1	12	12	20	15
Γ_+^1	1	1	1	1	1	1	1	1	1	1
Γ_+^3	3	τ	τ'	0	-1	3	τ	τ'	0	-1
$\Gamma_+^{3'}$	3	τ'	τ	0	-1	3	τ'	τ	0	-1
Γ_+^4	4	-1	-1	1	0	4	-1	-1	1	0
Γ_+^5	5	0	0	-1	1	5	0	0	-1	1
Γ_-^1	1	1	1	1	1	-1	-1	-1	-1	-1
Γ_-^3	3	τ	τ'	0	-1	-3	$-\tau$	$-\tau'$	0	1
$\Gamma_-^{3'}$	3	τ'	τ	0	-1	-3	$-\tau'$	$-\tau$	0	1
Γ_-^4	4	-1	-1	1	0	-4	1	1	-1	0
Γ_-^5	5	0	0	-1	1	-5	0	0	1	-1

Table 2: Characters of the irreducible representations of the icosahedral group H_3 and the subgroup $\mathcal{I}$ which does not include the inversion element (top left quadrant). The notation used is $\tau = \frac{1}{2}(1+\sqrt{5})$ and $\tau' = \frac{1}{2}(1-\sqrt{5})$, and $\mathcal{C}(g)$ denotes all elements in the conjugacy class of the element g.

References

[1] R. Twarock, "A tiling approach to virus capsid assembly explaining a structural puzzle in virology", *J. Theor. Biol.* **226** (2004) 477.

[2] J. E. Humphreys, "Reflection groups and Coxeter groups", vol. 29 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[3] D. Caspar and A. Klug, "Physical principles in the construction of regular viruses", *Cold Spring Harbor Sympos. Quant. Biol.* **27** (1962) 1–24.

[4] T. Keef, A. Taormina, and R. Twarock, "Assembly models of papovaviridae based on tiling theory", *Phys. Biol.* **2** (2005) 175–188, `q-bio/0508031`.

[5] R. Twarock, "Mathematical virology: a novel approach to the structure and assembly of viruses", *Philos. Trans. Roy. Soc.* **364** (2006) 3357–3373.

[6] K. M. ElSawy, A. Taormina, R. Twarock, and L. Vaughan, "Dynamical implications of viral tiling theory", *J. Theor. Biol.* **252** (2008) 357–369, `arXiv:0711.0541 [q-bio.BM]`.

[7] K. T. Tsen, S.-W. D. Tsen, O. F. Sankey, and J. G. Kiang, "Selective inactivation of micro-organisms with near-infrared femtosecond laser pulses", *J. Phys.: Cond. Mat.* **19** (2007) 472201.

[8] F. Tama and C. L. Brooks, "The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus", *J. Mol. Biol.* **318** (2002) 733.

[9] F. Tama and C. L. Brooks, "Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis", *J. Mol. Biol.* **345** (2005) 299.

[10] F. Tama and C. L. Brooks, "Symmetry, form, and shape: guiding principles for robustness in macromolecular machines", *Annu. Rev. Biophys. Biomol. Struct.* **35** (2006) 115–133.

[11] M. M. Gibbons and W. S. Klug, "Mechanical modeling of viral capsids", *J. Material Sci.* **42** (2007) 8995–9004.

[12] P. Durand, G. Trinquier, and Y.-H. Sanejouand, "A new approach for determining low-frequency normal modes in macromolecules", *Biophys.* **34** (1994) 759–771.

[13] A. D. Schuyler and G. S. Chirikjian, "Efficient determination of low-frequency normal modes of large protein structures by cluster-NMA", *J. Mol. Graphics and Modelling* **24** (2005) 46–58.

[14] I. Bahar and A. J. Rader, "Coarse-grained normal mode analysis in structural biology", *Curr. Opin. Struct. Biol.* **15** (2005) 586–592.

[15] C. Shepherd, I. Borelli, G. Lander, P. Natarajan, V. Siddavanahalli, C. Bajaj, J. Johnson, C. I. Brooks, and V. Reddy, "VIPERdb: a relational database for structural virology", *Nucl. Acids Res.* **34** (2006) D386–D389.

[16] H. W. T. v. Vlijmen and M. Karplus, "Normal mode analysis of large systems with icosahedral symmetry: Application to $(\text{Dialanine})_{60}$ in full and reduced basis set implementations", *J. Comput. Phys.* **115** (2001) 691–698.

[17] H. W. T. v. Vlijmen and M. Karplus, "Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry", *J. Mol. Biol.* **350** (2005) 528–542.

[18] F. Englert, K. Peeters, and A. Taormina, "The twenty-four near instabilities of Caspar-Klug viruses", `arxiv:0804.4275`, accepted for publication in *Phys. Rev. E*.

[19] J. Patera and R. Twarock, "Affine extensions of noncrystallographic Coxeter groups and quasicrystals", *J. Phys.* **A35** (2002) 1551–1574.

[20] T. Keef, A. Taormina, and R. Twarock, "Classification of capped tubular viral particles in the family of Papovaviridae", *J. Phys.: Cond. Mat.* **18** (2006) S375–S387, `q-bio.BM/0510028`.

[21] T. Keef, C. Micheletti, and R. Twarock, "Master equation approach to the assembly of viral capsids", *J. Theor. Biol.* **242** (2006) 713–721, `q-bio.BM/0508030`.

[22] T. Keef, R. Twarock, and K. Elsawy, "Blueprints for viral capsids in the family of polyomaviridae", *J. Theor. Biol.*, 2008 doi:10.1016/j.jtbi.2008.04.029.

[23] E. Wigner *Göttinger Nachrichten*, 1930 133.

[24] E. Wilson, "The normal modes and frequencies of vibration of the regular plane hexagon model of the benzene molecule", *Phys. Rev.* **45** (1934) 706–714.

[25] D. Bishop, "Group Theory and Chemistry", Clarendon Press, Oxford, 2003.

[26] J. A. McCammon, B. R. Gelin, M. Karplus, and P. Wolynes, "The hinge-bending mode in lysozyme", *Nature* **262** (1976) 325.

[27] T. Noguti and N. Go, "Collective variable description of small-amplitude conformational fluctuations in a globular protein", *Nature* **296** (1982) 776.

[28] B. R. Brooks and M. Karplus, "Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor", *Proc. Nat. Acad. Sci.* **80** (1983) 6571.

[29] N. Go, T. Noguti, and T. Nishikawa, "Dynamics of a small globular protein in terms of low-frequency vibrational modes", *Proc. Nat. Acad. Sci.* **80** (1983) 3696.

[30] M. Levitt, C. Sander, and P. Stern, "Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme", *Int. J. Quant. Chem.* **10** (1983) 181.

[31] R. Harrison, "Variational calculation of the normal modes of a large macromolecule: methods and some initial results", *Biopolymers* **23** (1984) 2943.

[32] T. Simonson and D. Perahia, "Normal modes of symmetric protein assemblies", *Biophys. J.* **61** (1992) 427.

[33] A. Rader, D. Vlad, and I. Bahar, "Maturation dynamics of bacteriophage HK97", *Structure* **13** (2005) 413–421.

[34] B. R. Brooks and M. Karplus, "Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme", *Proc. Nat. Acad. Sci.* **82** (1985) 4995.

[35] J. F. Gibrat and N. Go, "Normal mode analysis of human lysozyme: study of the relative motion of the two domains and characterization of the harmonic motion", *Proteins* **8** (1990) 258.

[36] O. Marques and Y. H. Sanejouand, "Hinge-bending motion in citrate synthase arising from normal mode calculations", *Proteins* **23** (1995) 557.

[37] L. Mouawad and D. Perahia, "Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures", *J. Mol. Biol.* **258** (1996) 393.

[38] V. Alexandrov, U. Lehnert, N. Echols, D. Milburn, D. Engelman, and M. Gerstein, "Normal modes for predicting protein motions: A comprehensive database assessment and associated Web tool", *Protein Sci.* **14** (2005) 633.

[39] M. M. Tirion, "Large amplitude elastic motions in proteins from a single-parameter, atomic analysis", *Phys. Rev. Lett.* **77** (1996) 1905–1908.

[40] J. I. Jeong, Y. Jang, and M. K. Kim, "A connection rule for α-carbon coarse-grained elastic network models using chemical bond information", *J. Mol. Graphics and Modelling* **24** (2006) 296.

[41] J. I. Jeong, Y. Jang, and M. K. Kim, "Using harmonic analysis and optimization to study macromolecular dynamics", *Int. J. Control, Automation and Systems* **4** (2006) 382.

[42] J. I. Jeong, Y. Jang, and M. K. Kim, "UMMS: constrained harmonic and anharmonic analyses of macromolecules based on elastic network models", *Nucl. Acids Res.* **34** (2006) W57.

[43] P. Freddolino, A. Arkhipov, S. Larson, A. McPherson, and K. Schulten, "Molecular dynamics simulations of the complete satellite tobacco mosaic virus", *Structure* **14** (2006) 437–449.

[44] B. Hespenheide, D. Jacobs, and M. Thorpe, "Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus", *J. Phys.: Cond. Mat.* **16** (2004) S5055–S5064.

[45] M. Kim, R. Jernigan, and G. Chirikjian, "An elastic network model of HK97 capsid maturation", *J. Struct. Biol.* **143** (2003) 107–117.

[46] W. K. Baskerville, "Vibrational spectrum of the $B = 7$ Skyrme soliton", `hep-th/9906063`.

[47] D. J. Jacobs and M. F. Thorpe, "Generic rigidity percolation: the pebble game", *Phys. Rev. Lett.* **75** (1995) 4051–4054.

[48] G. Laman, "On graphs and rigidity of plane skeletal structures", *J. Engineering Math.*, 1970 331–340.

[49] J. F. Cornwell, "Group theory in physics, vol. 1", Academic Press, 1984.

[50] Y.-H. Sanejouand, "Les modes normaux de vibration de basse fréquence des protéines", PhD thesis, Ecole Normale Supérieure Lyon, 2007.

[51] A. Atilgan, S. Durell, R. Jernigan, M. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model", *Biophys. J.* **80** (2001) 505–515.

[52] R. B. Hoyle, "Shapes and cycles arising at the steady bifurcation with icosahedral symmetry", *Physica* **D191** (2004) 261–281.